Large deviations for kernel density estimators and study for random decrement estimator

Liangzhen Lei

N^o d'ordre : D.U. 1624

UNIVERSITÉ BLAISE PASCAL

(U.F.R. de Recherche Scientifique et Technique)

ÉCOLE DOCTORALE DES SCIENCES FONDAMENTALES

N^o : 469

THÈSE

présentée pour obtenir le grade de

DOCTEUR D'UNIVERSITÉ

(Spécialité : MATHÉMATIQUES APPLIQUÉES)

par

Liangzhen LEI

GRANDES DÉVIATIONS POUR LES ESTIMATEURS À NOYAU DE LA DENSITÉ ET ÉTUDE DE L'ESTIMATEUR DE DÉCRÉMENT ALÉATOIRE

Soutenue publiquement le 9 décembre 2005, devant la commission d'examen :

Messieurs	Pierre BERNARD	Directeur de thèse
	Fabrice GAMBOA	Rapporteur
	Christian LÉONARD	Examinateur
	Jean PICARD	Examinateur
	Alain ROUAULT	Président, Rapporteur
	Liming WU	Directeur de thèse

Remerciements

Je tiens tout d'abord à remercier Liming WU pour l'encadrement de mes études de Master en Chine et de cette thèse de doctorat. Durant ces cinq années, il a été porteur d'idées et de soutiens permanents. Je remercie vivement Pierre BERNARD, pour sa co-direction de thèse. Il m'a guidée vers un domaine de recherches plus appliquées et m'a fortement encouragée pendant ces deux années d'études en France.

Je remercie très chaleureusement les rapporteurs de ma thèse Fabrice GAMBOA, rapporteur de thèse également en Chine, ainsi que Alain ROUAULT, pour leurs nombreuses suggestions constructives. Ils ont contribué à améliorer ce mémoire.

Mes remerciements s'adressent aussi à Jean PICARD et Christian LÉONARD qui me font l'honneur et le plaisir de participer au Jury.

Toute ma sympathie et reconnaissance à Erwan SAINT-LOUBERT-BIÉ pour son aide et ses pertinentes corrections sur l'introduction de ce travail ; tous mes sentiments d'amitié à Nöelle ROUGANNE pour ses encouragements et son soutien. Merci également à Sylvie PAYCHA, Yuejun PENG, Malek BOUAZIZ, Hacène DJELLOUT, Geneviève BERNARD, Marie-Paule BRESSOULALY pour leur attention.

Merci à Josephe BADRIKIAN et Paul-Louis HENNEQUIN pour leur présence, leur affection, comme mes grands parents en France.

Mathieu GOURCY, Ingrid VIOLET, les Docteurs Benoît TESTUD, Fabrice BLACHE, m'ont guidée attentivement dans les loisirs et dans l'apprentissage d'expressions françaises, très particulières, je leur témoigne toute mon amitié. J'ai été très heureuse de partager de bons moments avec Hamid BADI, Rémi SART, Malcom DJENNO, Béranger AUBIN, Vivien CLAUZON, Mamadou SY, Fana TANGARA, je les remercie pour leur bonne humeur.

Je n'oublie pas non plus tous les membres du département de Mathématiques, auxquels je témoigne ma profonde sympathie.

Que tous mes amis chinois, étudiants sur Clermont, doctorants, post-doctorants sachent que les bons moments partagés m'ont été précieux .

Mille merci aussi à mes copines Yutao MA, Xiaoqun ZHANG, Yuan YUAN.

À mes parents,

À ma soeur et sa famille.

Sommaire

Introduction et Résultats Principaux

Cette thèse de doctorat d'Université a été élaborée sous la direction des Professeurs Liming Wu et Pierre Bernard au sein du Laboratoire de Mathématiques de l'Université Blaise Pascal (Clermont-Ferrand II). Elle aborde des deux thèmes. Le premier est à l'étude systématique de principes de grandes déviations (abrégé en PGD dans la suite) pour l'estimateur à noyau de la densité des processus stochastiques. Le second est l'étude rigoureuse de l'estimateur de décrément aléatoire, qui est très souvent utilisé par les mécaniciens et les ingénieurs en raison de son efficacité numérique. Toutefois les comportements asymptotiques la consistance forte et TCL, sont mal compris jusqu'aujourd'hui.

Le contenu de ce travail sur le premier thème peut se résumer de la manière suivante. Le début est d'abord consacrée à des rudiments de la théorie de grandes déviations; puis nous présentons des rappels sur l'estimateur à noyau de la densité, avec la problématique qui nous occupe, c'est-à-dire, l'estimation de grandes déviations dans L^1 de l'estimateur à noyau de la densité dans le cas dépendant, ce qui est plus fort que le PGD de Donsker-Varadhan pour les mesures empiriques, et beaucoup plus difficile que dans le cas i.i.d.. Même dans le cas i.i.d., cependant les résultats connus sont très récents. Enfin, nous présentons les résultats obtenus dans les parties suivantes.

Les quatre premiers chapitres présentent quatre articles. Le premier concerne l'estimation de grandes déviations pour l'estimateur à noyau de la densié dans le cas i.i.d., et a été publié dans un proceeding [63]. Le deuxième a démontré la convergence exponentielle de cet estimateur, et une inégalité de concentration du type Hoeffding pour les processus ϕ-mélangeants, sous la condition très générale de la sommabilité des coefficients ϕ-mélangeants : il a été publié dans *Ann L'I.S.U.P.* [64]. Le troisième est paru dans *Stochastic Process. Appl.* [65], et démontre le PGD de l'estimateur de densité de la mesure invariante pour un processus de Markov uniformément ergodique, une hypothèse naturelle introduite par Deuschel et Stroock pour le PGD de loi mesure empirique. C'est un premier résultat dans cette direction pour le cas dépendant. Le quatrième article étudie les processus de Markov réversibles, pour lesquels, en général, l'hypothèse d'ergodicité uniforme de Deuschel-Stroock n'est pas satisfaite. Il démontre que tous les résultats de grandes déviations pour l'estimateur de densité dans le cas uniformément ergodique s'étendent à ce nouveau cadre sous l'hypothèse d'intégrabilité uniforme du noyau de transition dans L^2 (cette dernière hypothèse est une condition nécessaire et suffisante pour obtenir le PGD de mesures empiriques dans le cas réversible, voir [97]). Il est accepté pour publication par *Bernoulli*.

Mes travaux sur ce premier thème sont sous la direction du professeur Liming Wu.

Le deuxième thème de cette thèse est dans une ligne de recherches complètement différente et ses études sont réalisées sous la direction du professeur Pierre Bernard. On y étudie l'estimateur de décrément aléatoire (abrégé en EDA) pour des processus gaussiens stationnaires. Cette estimateur qui est couramment utilisé par les mécaniciens : (il est numériquement beaucoup plus rapide que l'estimateur classique). Cependant la vraie nature de l'EDA est mal comprise.

Dans le (dernier) chapitre 5 de cette thèse; on réalise l'analyse asymptotique de l'EDA. On y montre une loi forte des grands nombres, l'expression explicite du biais asymptotique de cet estimateur, ainsi que le Théorème de Limite Centrale (TCL). (Les résultats sont fondamentaux pour les inférences statistiques de cet estimateur). L'existence du biais asymptotique de l'EDA (pour la quelle on donne une condition necessaire et suffisante) offre un nouveau exemple illustrant le fameux "Paradoxe de Kac-Slepian" [55], et corrige une erreur courante dans la littérature des ingénieurs!

Partie I

0.1 Rudiments de la théorie des grandes déviations

0.1.1 Origines et définitions

La théorie moderne des grandes déviations découle de travaux fondamentaux de Donsker et Varadhan pour des grandes déviations de processus de Markov, et de Freidlin-Wentzell pour les perturbations aléatoires de systèmes dynamiques. Après trente ans de développement par nombre de chercheurs, elle est devenue une branche importante de la théorie des probabilités. Nous commençons par le sujet le plus classique en probabilités, c'est-à-dire le comportement de la moyenne empirique de variables aléatoires indépendantes et identiquement distribuées (v.a.i.i.d. en bref). Soit $(\xi_n, n \in \mathbb{N})$ une suite de v.a.i.i.d. à valeurs dans $\mathbb{R}^d$, définie sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Considérons la moyenne empirique,

$$\frac{S_n}{n} = \frac{1}{n}\sum_{k=0}^{n-1} \xi_k,\ n \geq 1$$

la vraie moyenne est dénotée par $m = \mathbb{E}(\xi_0)$. Par la loi des grandes nombres,

$$\mu_n := \mathbb{P}\left(\left|\frac{S_n}{n} - m\right| > \delta\right) \stackrel{n\to\infty}{\longrightarrow} 0,$$

L'événement $\left\{\left|\dfrac{S_n}{n} - m\right| > \delta\right\}$ décrit la déviation entre S_n/n et m. Un problème intéressant et important est l'estimation de la probabilité de cette déviation : on veut connaître la probabilité de déviation de manière précise. En général, il est difficile de donner une réponse exacte. Cependant, si la probabilité de déviation tend vers zero exponentiellement vite et si l'on s'intéresse seulement à la vitesse exponentielle, on entre dans le domaine des grandes déviations.

En général, le principe de grandes déviations (PGD) caractérise les comportements pour la limite extrême, lorsque $n \to \infty$, d'une famille de mesures de probabilité $\{\mu_n\}_{n\in\mathbb{N}}$ sur $(X, \mathcal{A})$ par une fonction de taux. La caractérisation se fait via les bornes aysmptotiques exponentielles inférieures et supérieures sur les probabilités que $\{\mu_n\}$ assigne aux sous-ensemble mesurables de X. Au long de ce travail, X est toujours un espace de Hausdorff régulier (où régulier signifie: $\forall x \in X$, il existe une base de voisinages fermés de x), $\mathcal{A}$ est une tribu fixée qui contient une base de voisinages ouverts et une autre base de voisinages fermés pour tout x dans X. Si X est un espace métrique, on adopte la convention $\mathcal{A} := \sigma(B(x,r); x \in X, r > 0)$, où $B(x,r) := \{y \in X; d(x,y) < r\}$.

Définition 0.1 *Une fonction de taux I est une fonction semi-continue inférieurement sur X à valeurs dans $[0, +\infty]$. On dit qu'une fonction de taux est inf-compact ou bonne, si $\forall L \geq 0, [I \leq L]$ est un sous-ensemble compact de X.*

Remarquons que si X est un espace métrique, la propriété de semi-continuité inférieure peut être vérifiée par les critères séquentiels. Ainsi I est semi-continue inférieurement si et seulement si $\liminf_{x_n \to x} I(x_n) \geq I(x)$ pour tous $x_n \to x, x \in X$. Une bonne fonction de taux I peut toujours atteindre sa borne inférieure sur un sous-ensemble fermé non-vide.

Définition 0.2 *Si $\{\mu_n\}$ est une famille de mesures de probabilité sur X, on dit qu'elle satifait le Principe de Grandes Déviations (PGD) pour une fonction de taux I et de vitesse n, si I est une bonne fonction et si $\{\mu_n\}$ satisfait en même temps la borne inférieure des grandes déviations et la borne supérieure des grandes déviations , où*

(a) la borne inféreure de grandes déviations (LLD) : pour tout ensemble ouvert $G \in \mathcal{A}$,

$$l(G) := \liminf_{n \to +\infty} \frac{1}{n} \log \mu_n(G) \geq -\inf_G I$$

(b)la borne supérieure de grandes déviations (ULD) : pour tout ensemble fermé $F \in \mathcal{A}$,

$$u(F) := \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(F) \leq -\inf_F I$$

Dans les références usuelles, le principe faible de grandes déviations est aussi étudié ; il est inspiré par une approche naturelle : montrer d'abord la borne supérieure de grandes déviations pour les ensembles compacts.

Définition 0.3 *La famille $\{\mu_n\}$ satisfait le principe faible de grandes déviations (w-PGD), si (LLD) et (w-ULD) sont vérifiées , où w-ULD signifie :*

(w-ULD) pour tout emsemble compact $K \subset X$,

$$u(K) \leq -\inf_K I$$

Un autre type de PGD faible est considéré, c'est-à-dire, le principe faible* de grandes déviations suivant (w*-PGD) :

Définition 0.4 *La famille $\{\mu_n\}$ satisfait le principle faible* de grandes déviations (w*-PGD), si LLD et w*-ULD sont vérifiés, où w*-ULD signifie :*

(w-ULD) pour tout sous-ensemble compact $K \subset X$, pour $\delta > 0$ quelconque, il existe un ensemble ouvert et $\mathcal{A}$-mesurable $G^\delta \supset K$, tel que*

$$u(G^\delta) = \begin{cases} -\inf_K I + \delta & \text{si } \inf_K I < \infty \\ -\frac{1}{\delta} & \text{sinon.} \end{cases}$$

Proposition 0.5 *La famille (μ_n) satisfait le w^*-PGD de fonction de taux I sur un espace métrique (X, d), si et seulement si l'estimation locale suivante est vérifiée: pour tout $x \in X$,*

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(B(x,\delta)) = \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(B(x,\delta)) = -I(x)$$

où $B(x,\delta) = \{y \in X; d(x,y) < \delta\}$.

Il est facile de voir que w*-ULD $\Longrightarrow$ w-ULD et qu'ils deviennent équivalents si X est localement compact. La question "comment passer de w*-PGD à PGD" est discutée plus loin dans ce travail.

Pour une suite de variables aléatoires $(Z_n)_{n\in\mathbb{N}}$ à valeurs dans $(X, \mathcal{A})$ et définie sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, on dit que (Z_n) satisfait le PGD (resp. w^*-PGD), si c'est le cas pour la famille de probabilités $(\mu_n := \mathbb{P}(Z_n \in \cdot))$.

0.1.2 Fonctionnelle de Cramér et théorème de Gärtner-Ellis

Une methode classique d'étude des PGD est donnée par Cramér [21], qui a introduit le changement de mesure. Une extension aux distributions générales a été réalisée par Chernoff [17], qui a introduit la borne supérieure.

Le point crucial de leur méthode est de considérer la fonctionnelle de Cramér. Expliquons-la dans le cas classique suivant. Considérons les moyennes empiriques $S_n =: \dfrac{1}{n}\sum_{j=1}^n X_j$, pour les vecteurs aléatoires i.i.d. $X_1, \cdots, X_n, \cdots$, et à valeurs dans $\mathbb{R}^d$, avec X_1 ayant pour loi $\mu \in M_1(\mathbb{R}^d)$ (Où $M_1(\mathbb{R}^d)$ est l'espace des mesure de probabilité sur $\mathbb{R}^d$).

Définition 0.6 *Le logarithme de la fonction génératrice des moments associée à la loi μ est défini par*

$$\Lambda(\lambda) =: \log M(\lambda) =: \log \mathbb{E}[e^{\langle \lambda, X_1 \rangle}], \ \ \lambda \in \mathbb{R}^d$$

où $\langle \lambda, x \rangle := \sum_{j=1}^d \lambda_j x_j, \ , x \in \mathbb{R}^d$ est le produit scalaire dans $\mathbb{R}^d$, et x_j la j-ième coordonnée de x.

Définition 0.7 *La transformée de Fenchel-Legendre de* $\Lambda(\lambda)$ *est*

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, x \rangle - \Lambda(\lambda)\}, \ \ x \in \mathbb{R}^d$$

Dans le cas i.i.d., le théorème classique de Cramér dit

Théorème 0.8 *Si* $\Lambda(\lambda) < \infty$ *pour tous* λ*, alors* $\left\{\mu_n = \mathbb{P}\left(\frac{S_n}{n} \in \cdot\right), n \to \infty\right\}$ *satisfait le PGD sur* $\mathbb{R}^d$ *de bonne fonction de taux convexe* $\Lambda^*(\cdot)$*.*

En fait, la condition dans le théorème peut être affaiblie en :

(i) il existe $\delta > 0$*, tel que* $\Lambda(\lambda) < \infty$ *pour tous* $\lambda : |\lambda| \le \delta$*.*

De plus, Bahadur et Zabell [3] ont montré que $(\mu_n, n \to \infty)$ satisfait toujours le w-PGD sur $\mathbb{R}^d$ de fonction de taux $\Lambda^*(\cdot)$, sans aucune condition d'intégrabilité.

Maintenant, présentons le théorème de Gärtner-Ellis, un des outils les plus puissants dans la théorie des grandes déviations. Considérons une suite (Z_n) de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$. Posons

$$\Lambda_n(\lambda) =: \frac{1}{n} \log \mathbb{E}[e^{n\langle \lambda, Z_n \rangle}], \ \ \lambda \in \mathbb{R}^d$$

Hypothèse: Pour chaque $\lambda \in \mathbb{R}^d$,

$$\Lambda(\lambda) =: \lim_{n\to\infty} \Lambda_n(\lambda) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[e^{n\langle \lambda, Z_n \rangle}] \in (-\infty, +\infty]$$

existe (Λ va être appelée fonctionnelle de Cramér) et $\overset{\circ}{\mathcal{D}}_\Lambda = (\overset{\circ}{Dom}\Lambda) = [\Lambda \overset{\circ}{<} +\infty]$ (l'intérieur) n'est pas vide. De plus, $0 \in \overset{\circ}{\mathcal{D}}_\Lambda$.

Définition 0.9 *Une fonction convexe* Λ*:* $\mathbb{R}^d \to (-\infty, +\infty]$ *est essentiellement lisse si :*

(a)$\overset{\circ}{\mathcal{D}}_\Lambda$ *est non vide;*

(b)$\Lambda(\cdot)$ *est dérivable sur* $\overset{\circ}{\mathcal{D}}_\Lambda$*;*

(c)$\Lambda(\cdot)$ *est escarpée, c'est-à-dire ,* $\lim_{n\to\infty} \|\nabla\Lambda(\lambda_n)\| = \infty$*, où* $\{\lambda_n\}$ *est une suite quelconque dans* $\mathcal{D}_\Lambda^\circ$ *convergeant vers un point de la frontière de* $\overset{\circ}{\mathcal{D}}_\Lambda$*.*

Remarque: Si $\mathcal{D}_\Lambda = \mathbb{R}^d$ et $\Lambda(\cdot)$ est Gâteaux-différentiable, alors Λ est essentiellement lisse.

Le théorème suivant est un outil crucial dans la théorie des grandes déviations, Voir Dembo et Zeitouni ([25], Théorème 2.3.6. p44) ou Wu ([93], Théorème 1.4, p276).

Théorème 0.10 *(Gärtner-Ellis) Supposons que* **Hypothèse** *est vérifiée et que* μ_n *est la loi de* Z_n. *Alors*

(a) Pour tout ensemble fermé F,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \leq - \inf_{x\in F} \Lambda^*(x)$$

(b) Pour tout ensemble ouvert G,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geq - \inf_{x\in G\cap C_s(\Lambda^*)} \Lambda^*(x)$$

où $C_s(\Lambda^*)$ *est l'ensemble des points de stricte convexité de la fonction convexe* Λ^*, *c'est-à-dire,* $x_0 \in C_s(\Lambda^*)$ *si et seulement si* $\exists y_0 \in \mathcal{D}^0_\Lambda$, *tel que*

$$\forall x \in \mathbb{R}^d, x \neq x_0, \Lambda^*(x) > \Lambda^*(x_0) + \langle x - x_0, y_0\rangle.$$

(c) Si Λ *est essentiellement lisse, alors on a le PGD de fonction de taux* $\Lambda^*(\cdot)$.

Comme conséquence, on a

Théorème 0.11 *Supposons que* **Hypothèse** *est vérifiée , et que si* Λ *est Gâteaux-différentiable sur* $\mathbb{R}^d$, *alors on a le PGD de fonction de taux* $\Lambda^*(\cdot)$.

0.1.3 Du w*-PGD au PGD : exp-tension* et Le principe de contraction

Par les définitions de PGD et w*-PGD, on sait que le w*-PGD donne une estimation de voisinage locale, mais le PGD donne une estimation globale, donc passer de w*-PGD à PGD est une question bien naturelle, et le concept suivant est crucial :

Définition 0.12 *(Wu, [93], p246) Une suite de mesures de probabilité* $\{\mu_n, n \to \infty\}$ *est appelée* ***exp-tendue****, *si pour tout* $L > 0$, *il existe un ensemble compact* K_L, *tel que pour tout ensemble ouvert* $G \supset K_L$, $G \in \mathcal{A}$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(G^c) < -L.$$

On a le théorème suivant, dû à Wu:

Théorème 0.13 *(Wu, [93]), p246) Si $\{\mu_n, n \to \infty\}$ satisfait le w*-PGD de fonction de taux I sur X, alors $\{\mu_n, n \to \infty\}$ satisfait le PGD de même fonction de taux si et seulement si $\{\mu_n, n \to \infty\}$ est exp-tendue*.*

Remarques:

(a) Si $\{\mu_n, n \to \infty\}$ satisfait w*-ULD et est exp-tendue*, si I est inf-compacte, alors on a l'ULD.

(b) Si $\{\mu_n, n \to \infty\}$ satisfait un ULD, alors $\{\mu_n, n \to +\infty\}$ est exp-tendue*.

Le principe de contraction signifie la stabilité du PGD par une application continue.

Proposition 0.14 *(Principe de contraction) Soit Y est un espace de Hausdorff régulier, f est une application continue de X dans Y, $\mathcal{B} = \{B \subset Y | f^{-1}(B) \in \mathcal{A}\}$, $\nu_n(B) := \mu_n(f^{-1}(B))$, $\forall B \in \mathcal{B}$. Si $\{\mu_n, n \to \infty\}$ satisfait un PGD de fonction de taux I sur X, alors $\{\nu_n, n \to \infty\}$ satisfait un PGD sur Y de fonction de taux :*

$$I^f(y) = \inf\{I(x) | x \in X; f(x) = y\}, \forall y \in Y.$$

0.1.4 Références sur les applications des grandes déviations

Les techniques de grandes déviation ont été très utilisées en statistiques : ainsi dans les estimations des U-statistiques, des t-statistiques de Student, des statistiques auto-normalisées, dan les applications aux problèmes de tests d'hypothèse, les tests généralisés de rapports de vraisemblance, etc. Les cours font référence à Dembo et Zeitouni [25], Zeitouni et Gutman [99], Zeitouni et Zakai [100], Gamboa et Gassiat [38], Hoeffding [46], Watanabe et al. [90], Nakayama et al. [72], J.A. Bucklew [12, 13], Aleshkyavichene [2], Horváth et Shao [49], Shao [84, 85], Dembo et Shao [23, 24], Jing et al. [53], Djellout et al. [30] et aux réferences ces qu'on peut y obtenir.

Il y a aussi des applications en mécanique statistique et pour les systèmes de particules en interaction. Pour une introduction au PGD en mécanique statistique classique et pour les modèles de spin, voir Ellis [36], et pour des publications plus récentes, voir Kusuoka et Tanura [60], Durrett [34], Chaganty et Sethuraman [16], Donsker et Varadhan [32], Leonard [66], Bramson et al. [10], Durrett et Schonmann [35], Lebowitz et Schonmann [61], Orey [74], Deuschel [26], Donsker et Varadhan [32], Ben Arous et Brunaud [4], Cox et Durrett [20], Kipnis et Olla [57], Papangelou [75], Stroock et Zegarlinski [89], Ben Arous et Guionnet [5], Schonmann et Shlosman [87] et les références citées dans ces travaux. Une application particulièrement intéressante des techniques de grande déviations est la construction de grandes déviations plus précises dans les surfaces; voir, par exemple, Schonmann [86], Pfister [78], Dobrushin et al. [31], Ioffe [52], Pisztora [79].

0.2 Estimateur à noyau de la densité

0.2.1 Motivation

Il y a quarante ans, Parzen [76] a étudié les propriétés fondamentales de l'estimateur à noyau de la densité, juste après leur introduction par Rosenblatt [88]. A partir de ce moment là, cet estimateur à noyau de la densité est devenu un objet classique étudié par des statisticiens et des probabilistes. Pour les statisticiens, il est déjà devenu un exemple canonique d'estimateur non-paramétrique de courbe, qui a utilisé beaucoup d'idées importantes de la théorie d'approximation et l'analyse harmonique aux statistiques non-paramétriques. En fait, ils n'imposent pas de restriction paramétrique sur la forme fonctionnelle de la densité. Dans la suite, on explique la raison pour laquelle l'estimateur paramétrique ne marche pas certaines fois.

Le cadre usuel est le suivant (pour plus de pécision, voir [59]) :

Etant donné un échantillon i.i.d $X_i \sim X, i = 1, \cdots, n$, on s'intéresse à l'estimation de la densité de la distribution f.

Dans le cadre paramétrique, on va spécifier une classe de densités paramétr'ees par certains vecteurs de dimension finie $\theta \in \Theta \subseteq \mathbb{R}^d$, $\{f(\cdot;\theta)|\theta \in \Theta\}$, et puis supposer que la vraie densité est $f = f(\cdot;\theta_0)$ pour un certain $\theta_0 \in \Theta$. Un exemple standard de cette démarche est le cas de la loi normale où

$$f(x,\theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\{\frac{(x-u)^2}{2\sigma^2}\}$$

et $\theta = (u, \sigma^2)$. Un estimateur évident de la densité doit alors être $\hat{f} = f(\cdot, \hat{\theta})$, où

$$\hat{\theta} = \arg\min_{\theta\in\Theta} -\frac{1}{n}\sum_{i=1}^{n}\log f(X_i;\theta).$$

Sous les conditions régulières,

$$\sqrt{n}(\hat{f}(x) - f(x)) \xrightarrow{d} N(0, V_0(x)H_0^{-1}V_0(x)),$$

où $V_0(x) = \partial_\theta f(x;\theta)$, $H_0 = -\mathbb{E}[\partial_{\theta\theta}\log f(X;\theta_0)]$. Donc si on suppose que f appartient à la classe paramétrique spécifiée, on a un estimateur naturel qui marche bien.

Cependant, il y a un risque de fausse-spécification. (Voir [59] pour des détails). Supposons que $f \notin \{f(\cdot;\theta)|\theta \in \Theta\}$. Alors, l'estimateur paramétrique peut être biaisé et on a

$$\sqrt{n}(\hat{f}(x) - f(x;\bar{\theta})) \xrightarrow{d} N(0, \bar{V}(x)\bar{H}^{-1}\bar{G}\bar{H}^{-1}\bar{V}(x)),$$

où $\bar{\theta} = \arg\min_{\theta\in\Theta} \int_{\mathbb{R}^d} \log f(x;\theta) f(x)dx$ avec $f(x;\bar{\theta}) \neq f(x)$, et

$$\bar{G} = \mathbb{E}[\partial_\theta \log f(X;\bar{\theta})\partial_\theta \log f(X;\bar{\theta})]$$

$$\bar{H} = -\mathbb{E}[\partial_{\theta\theta} \log f(X;\bar{\theta})], \bar{V}(x) = \partial_\theta f(x;\bar{\theta}).$$

Une solution à ce problème est de choisir une classe paramétrique plus grande et plus flexible. Toutefois aussi grande soit la classe choisie, on ne peut jamais éviter complètement le risque que $f \notin \{f(\cdot;\theta)|\theta \in \Theta\}$. En outre, les problèmes numériques associés pour obtenir effectivement $\hat{\theta}$ croissent avec la taille de l'ensemble des paramètres.

Une autre solution est de construire un estimateur non-paramétrique qui marche pour toutes les densités, sans imposer aucune forme paramétrique sur. Cela peut enlever tous les risques de fausse-spécification. Un exemple d'un tel estimateur non-paramétrique de densité, est l'estimateur à noyau de la densité .

Quand on regarde un phénomène naturel évoluant avec le temps, l'échantillon obtenu n'est pas toujours i.i.d., mais souvent Markovian. Les processus de Markov sont la classe la plus importante de processus stochastiques à la fois pour la théorie et les applications. Ainsi il constitue aussi un modèle basique en statistique. Il est donc important d'estimer leur mesures invariantes et leurs probabilités de transition ou leurs densités. C'est exactement ce que nous allons faire à présent, dans le cas dépendant.

0.2.2 Définition et références

Soit $\{X_i, i \geq 1\}$ une suite de variables aléatoires stationnaires et ergodiques à valeurs dans $\mathbb{R}^d$, définie sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, possédant une distribution commun $d\mu = f(x)dx$, la densité $f \in L^1(\mathbb{R}^d)$ étant inconnue. La mesure empirique est définie comme $L_n = \dfrac{1}{n}\sum_{i=1}^n \delta_{X_i}$. Soit K une densité de probabilité sur $\mathbb{R}^d$, c'est-à-dire :

(H1) $K \geq 0$, *et* $\int_{\mathbb{R}^d} K dx = 1$,

et soit $K_h(x) = \dfrac{1}{h^d} K(\dfrac{x}{h})$. L'estimateur à noyau de la densité de f est défini suivant:

$$f_n^*(x) = K_{h_n} * L_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h_n^d} K\left(\frac{x - X_i}{h_n}\right), \quad x \in \mathbb{R}^d \tag{0.1}$$

où $h = h_n, \{h_n, n \geq 1\}$ est une suite de nombres strictement positifs (h_n est appelée la fenêtre) satisfaisant

(H2) $h_n \to 0, \quad nh_n^d \to +\infty \quad lorsque \quad n \to \infty.$

Une distance naturelle et très utilisée entre f_n^* et la densité inconnue f est la distance dans L^1 :

$$D_n^* = \int_{\mathbb{R}^d} |f_n^*(x) - f(x)| dx. \tag{0.2}$$

Nous étudions les grandes déviations pour f_n^*.

Resultats connus dans le cas i.i.d.

Dans le cas i.i.d., Devroye [28] a montré que tous les types de consistance dans $L^1(\mathbb{R}^d)$ de f_n^* sont équivalents à l'hypothèse (**H2**) sur la fenêtre. La normalité asymptotique de D_n^* a été étudiée par Csörgö et Horváth [22]. En dimension $d = 1$, Louani a montré dans [67] un PGD pour $f_n^*(t)$, de vitesse nh_n, pour tout point quelconque $t \in \mathbb{R}$ et dans [68] le PGD pour D_n^* de vitesse n. Joutard [54] a établi le même PGD pour f_n^* mais dans des hypothèses différentes. Plus récemment, Gao a obtenu dans[39] le PGD et le principe de déviation modérée (PDM) de f_n^* dans $L^\infty(\mathbb{R}^d)$, pour une vitesse dependant de la fenêtre h_n, avec une function de taux dependant de K. Ce sont, à notre connaissance, les seuls résultats de grandes déviations pour l'estimateur à noyau de la densité.

- `PGD ponctuel et uniforme de` $f_n^*(x)$, $x \in \mathbb{R}$ `(Louani, [67])`

Pour prouver son résultat de grande déviations, Louani a introduit les hypothèses suivantes:

(LOU1) La fonction f est dérivable sur $\mathbb{R}$, de dérivée f' et $\sup_{x \in \mathbb{R}} \|f'(x)\| < \infty$.

(LOU2) La fonction

$$I(t) = \int_{\mathbb{R}} (\exp(tK(z)) - 1) dz$$

est finie pour tout $t > 0$. De plus elle est deux fois dérivable sur $\mathbb{R}$.

(LOU3) Pout tout $t > 0$,

$$\int_{\mathbb{R}} \|z\| (\exp(tK(z)) - 1) dz < +\infty.$$

Sou ces trois hypothèses, il a obtenu le résultat suivant :

Théorème 0.15 (theorem 1, [67]) *Supposons que les hypothèses* **(H1)** *et* **(H2)** *sont vérifiées et que les hypothèses (LOU1), (LOU2) et (LOU3) sont satisfaites. Alors,*

(1) Pour tout $x \in \mathbb{R}$ *et tout* $\alpha > 0$,

$$\lim_{n\to\infty} \frac{1}{nh_n} \log \mathbb{P}(f_n^*(x) - f(x) > \alpha) = -\Lambda^*(\alpha)$$

où $\Lambda^*(\alpha) = \sup_{t>0}\{t(\alpha + f(x)) - f(x)I(t)\}$

(2) Pour tout $x \in \mathbb{R}$ *et tout* $\alpha > 0$,

$$\lim_{n\to\infty} \frac{1}{nh_n} \log \mathbb{P}(f_n^*(x) - f(x) < -\alpha) = -\Lambda^*(-\alpha);$$

(3) Pour tout $x \in \mathbb{R}$ *et tout* $\alpha > 0$,

$$\lim_{n\to\infty} \frac{1}{nh_n} \log \mathbb{P}(\|f_n^*(x) - f(x)\| > \alpha) = -\Lambda^*(\alpha).$$

Il a également donné des résultats de grandes déviations pour la déviation uniforme, c'est-à-dire, pour $\|f_n^*(x) - f(x)\|_\infty := \sup_{x\in\mathbb{R}} |f_n^*(x) - f(x)|$.

Théorème 0.16 (theorem 2, [67]) *Supposons que les hypothèses* **(H1)** *et* **(H2)** *sont vérifiées et que les hypothèses (LOU1) , (LOU2) et (LOU3) sont satisfaites. Supposons de plus :*

(LOU4) K *est à support borné et Lipschitzienne continue,*

(LOU5) il existe une suite de nombres positifs $\{H_n, n \geq 1\}$ *tendant vers l'infinie telle que* $\lim_{n\to\infty} \log(H_n\alpha_n^{-2})/n\alpha_n = 0$ *et, pour un* $\tau > 0$,

$$\lim_{n\to\infty} \frac{\mathbb{P}(\|X_1\| > H_n)\exp\{\tau n\alpha_n\}}{n\alpha_n^2} = 0$$

Alors,

$$\lim_{n\to\infty} \frac{1}{n\alpha_n} \log \mathbb{P}(\|f_n^*(x) - f(x)\|_\infty > \lambda) = -g(\lambda)$$

- `PGD de` D_n^* `(Louani, [68])`

Avant présenter son résultat, définissons les fonctions suivantes: pour tous $0 \leq a \leq 1$,

$$\Gamma_a^+(r) := \begin{cases} (a+\frac{r}{2})\log(1+\frac{r}{2a}) + (1-a-\frac{r}{2})\log(1-\frac{r}{2(1-a)}) & \text{si } \ 0 < r < 2-2a \\ +\infty, & \text{sinon} \end{cases} \tag{0.3}$$

$$\Gamma_a^-(r) := \begin{cases} (a-\frac{r}{2})\log(1-\frac{r}{2a}) + (1-a+\frac{r}{2})\log(1+\frac{r}{2(1-a)}) & \text{si } \ 0 < r < 2a \\ +\infty, & \text{sinon} \end{cases} \tag{0.4}$$

$$\begin{aligned} \Gamma_a(r) &:= \min\{\Gamma_a^+(r), \Gamma_a^-(r)\}; \\ l(r) &:= \inf\{\Gamma_a(r) : 0 \leq a \leq 1\} \end{aligned} \tag{0.5}$$

Maintenant, le résultat principal de Louani [68] est

Théorème 0.17 (Louani) *Supposons que les hypothèses* **(H1)** *et* **(H2)** *sont vérifiées, alors D_n^* satisfait le PGD de fonction de taux $l(\cdot)$, c'est-à-dire :*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\|f_n^* - f\|_1 > r\right) = -l(r), \ \forall r > 0. \tag{0.6}$$

- `PGD de` f_n^* `(Joutard, [54])`

Joutard a considéré le même cas que Louani dans sa thèse. Les hypothèses sur K et f sont suivantes (comparer avec celles de Louani)

(JOU1) f est continue sur $\mathbb{R}$ et $\|f\|_\infty = \sup_{y\in\mathbb{R}} |f(y)| < \infty$

(JOU2) $\phi_0(t) = \int_{\mathbb{R}} K(z)\exp\left(tK(z)\right)dz$ est finie pour tout $t \in (-\infty, a], a > 0$

Sous ces hypothèses, elle a obtenu le résultat suivant :

Théorème 0.18 (théorème 3.2.1., [54]) *Supposons (JOU1) et (JOU2) vérifiées et $\lim_{n\to\infty} nh_n = +\infty$. Alors, pour tout $x \in \mathbb{R}$ tel que $f(x) > 0$, $f_n^*(x) - f(x)$ satisfait un PGD de vitesse (nh_n), et de bonne fonction de taux*

$$\Lambda^*(u) = \sup_{t\in\mathbb{R}}\{t(u + f(x)) - f(x)\psi(t)\},$$

où $\psi(t) = \int_{\mathbb{R}}(\exp(tK(z)) - 1)dz$.

Pour cette preuve, puisque les hypothèses considérées ne lui permettent plus d'appliquer le théorème de Gärtner-Ellis (0.10), elle a utilisé un changement de probabilité exponentiel.

- `PGD pour` f_n^* `dans` $L^\infty(\mathbb{R}^d)$ `(Gao, [39])`

Gao a montré que si la fonction à noyau etait une fonction intégrable à variation bornée, et que la densité commune f des variables aléatoires était continue , avec $f(x) \to 0$ lorsque $|x| \to \infty$, alors, on avait le PGD et le PDM pour $\|f_n^*(x) - \mathbb{E}(f_n^*(x))\|_\infty := \{\sup_{x\in\mathbb{R}^d} |f_n^*(x) - \mathbb{E}f_n^*(x)|,\ n \geq 1\}$. Tout d'abord, présentons les hypothèses dans le théorème de PGD.

Hypothèses:

(GAO1) $\{h_n, n \geq 1\}$ est la fenêtre, qui est une suite de nombres strictement positifs satisfaisant :

$$h_n \to 0,\ \ nh_n^d \to +\infty,\ \ \frac{nh_n^d}{\log {h_n}^{-1}} \longrightarrow 0, \quad lorsque n \to \infty$$

(GAO2) f est continue et

$$\lim_{x\to\infty} f(x) = 0$$

(GAO3) K est une fonction bornée, de carré intégrable et appartient à l'espace vectoriel engendré par les fonctions $k \geq 0$ qui satisfont la propriété suivante: l'hypographe de k, $\{(s,u); k(s) \geq g(u)\}$, peut être écrit sous la forme d'un nombre fini d'intersections ou d'unions d'ensembles du type $\{(s,u); p(s,u) \geq \psi(u)\}$, où p est un polynôme sur $\mathbb{R}^d \times \mathbb{R}$ et ψ is une fonction réelle arbitraire. On suppose en outre que K est intégrable: $\int_{\mathbb{R}^d} |K(x)dx| < \infty$.

Théorème 0.19 (Gao) *Supposons que K est positive et que les hypothèses (GAO1), (GAO3), (GAO3) sont vérifiées. Alors pour tout $\lambda > 0$,*

$$\lim_{n\to\infty} \frac{1}{nh_n^d} \log \mathbb{P}(\|f_n^* - \mathbb{E}f_n^*\|_\infty > \lambda) = -J(\lambda);$$

et

$$\lim_{n\to\infty} \frac{1}{nh_n^d} \log \mathbb{P}(\|f_n^* - \|K\|_1 f\|_\infty > \lambda) = -J(\lambda);$$

où

$$J(\lambda) = \inf_{x\in\mathbb{R}^d} \sup_{t\in\mathbb{R}} \{t\lambda - f(x) \int_{\mathbb{R}^d} (\exp\{tK(x)\} - 1 - tK(x))dx\}.$$

En particulier, si $\int_{\mathbb{R}^d} K(x) = 1$*, alors pour tout* $\lambda > 0$*,*

$$\lim_{n\to\infty} \frac{1}{nh_n^d} \log \mathbb{P}(\|f_n^* - f\|_\infty > \lambda) = -J(\lambda).$$

Les autres théorèmes limites pour l'estimateur à noyau de la densité dans le cas i.i.d., comme le principe de déviation modérée (PDM), la loi du logarithme itéré (LLI), le théorème de la limite centrale (TCL), sont aussi étudiés activement. Par exemple, Gao [40] obtient le PGD pour f_n^* dans $L^1(\mathbb{R}^d)$ et la LLI pour D_n^*. Giné et al. [42] obtiennent un TCL fonctionnelle et un théorème du type de Glivenko-Cantelli pour le processus d'estimateur de densité, sous la norme de L^1. Les résultats intéressants les plus récents, sur le comportement de la limite de f_n^* dans le cas i.i.d. ont été obtenus également par Giné et al. [44], Giné et Mason [43]. Ainsi, dans [44] , ils obtiennent la LLI pour la déviation absolue p-ième intégrable entre l'estimateur à noyau de la densité et sa moyenne. Notons aussi que Menneteau [70] a étudié des grandes déviations uniformes pour des processus empiriques locaux. Ses résultats peuvent être utilisés pour retrouver ceux de Louani.

Dans le cas dépendant

Dans le cas dépendant, l'estimateur à noyau de la densité est également étudié par de nombreux chercheurs, mais ces études sont surtout concentrées sur la consistance et le théorème de limite centrale, voir Peligrad [77], Adams et Nobel [1], Bosq et al. [9] et les références citées dans leurs travaux. Mais on sait très peu de choses sur des grandes déviations et des déviations modérées.

L'extension des résultats de grandes déviations et de déviations modérées du cas i.i.d. ([68], [39], [40]) au cas dépendant est une question intéressante et ouverte. Elle est le problème principal dans cette thèse. Cette question est très delicate car même pour les chaînes de Markov stationnaires et récurrentes de Doeblin, le PGD pour les mesures empiriques peut ne pas exister en général, voir Bryc et Dembo [11].

Dans la thèse de Worms [91], sous l'hypothèse d'un PGD supérieur pour les mesures empiriques, on établit un PGD ponctuel et un PGD uniforme sur les compacts(et des bornes de type Chernov pour la topologie de la norme uniforme) pour un estimateur à noyau de la densité de la loi stationnaire d'une chaîne de Markov stable. Ce resultat généralise ceux de D. Louani dans [67], qui concernaient l'estimation de la densité d'une suite i.i.d..

Ses hypothèses sur la chaîne de Markov sont les suivantes : Soit $(X_n)_{n\geq 0}$ une chaîne de Markov à valeurs dans $\mathbb{R}^d$, dont la probabilité de transition satisfait aux conditions qui suivent :

(WORMS1) pout tout $x \in \mathbb{R}^d$, $\pi(x, dy)$ admet une densité $p(x, \cdot)$ par rapport à la mesure de Lebesgue :

$$\pi(x, dy) = p(x, y)dy$$

Cette densité p est supposée unifomément bornée (par $\|f\|_\infty$) et lipschitzienne en chacune des deux variables : on suppose donc qu'il existe des constantes α_1 et α_2 telles que, pour tous x, x', y, y' dans $\mathbb{R}^d$

$$|p(x, y) - p(x, y')| \leq \alpha_1 \|y - y'\|$$

$$|p(x, y) - p(x', y)| \leq \alpha_2 \|x - x'\|$$

(ainsi π est fellérienne).

(WORMS2) la chaîne admet une unique loi invariante μ;

(WORMS3) la suite des mesures empiriques $L_n(\cdot) = \frac{1}{n} \sum_{j=1}^n \delta_{X_{j-1}}(\cdot)$ satisfait à un PGD supérieur dans l'espace $M_1(\mathbb{R}^d)$ muni de la convergence étroite.

La loi invariante μ admet une densité $f(\cdot)$ relativement à la mesure de Lebesgue: puisque $f(y) = \int p(x, y)\mu(dx)$, et $f(\cdot)$ est bornée et lipschitzienne de coefficient de Lipschitz α_1.

Il a les PGD suivants (ainsi que les bornes de Chernov uniformes).

Théorème 0.20 (Worms, [91]) *On considère un noyau K et une chaîne de Markov satisfaisant aux hypothèses (WORM1), (WORM2), (WORM3),* **(H1)** *et* **(H2)**.

(a) (PGD) Pour chaque $x \in \mathbb{R}^d$ et toute suite (x_n) convergeant vers x, la suite $(f_n^*(x_n))$ satisfait un PGD de vitesse (nh_n^d) et de taux $f(x)I(\frac{\cdot}{f(x)})$, où

$$I(t) = \sup_{s \in \mathbb{R}} \left(st - \int (e^{sK(z)} - 1)dz \right)$$

En particulier pour chaque $x \in \mathbb{R}^d$, on a $f_n^*(x) \overset{n \to \infty}{\longrightarrow} f(x)$ exponentiellement vite.

(b)(PGD uniforme sur les compacts) Pour tout compact $H \subset \{f > 0\}$,

$$\liminf \inf_{x \in H} \left\{ \frac{1}{nh_n^d} \log \mathbb{P}[f_n^*(x)/f(x) \in U] + f(x)I(U) \right\} \geq 0$$

$$\limsup \sup_{x \in H} \left\{ \frac{1}{nh_n^d} \log \mathbb{P}[f_n^*(x)/f(x) \in F] + f(x)I(F) \right\} \leq 0$$

pour tout ouvert U de $\mathbb{R}$ et tout fermé F de $\mathbb{R}$.

Ce résultat de forme ponctuelle a un caractère étrange. La fonction de taux reste la même que celle du cas i.i.d. et on n'y voit pas apparaître la structure dépendante du processus. L'extension des résultats en forme fonctionnelle de Joutard et de Gao, au cas dépendant reste une question ouverte.

Les études de grandes déviations pour les mesures d'occupation de processus de Markov ont été initiées par Donsker-Varadhan [32]. Ils ont montré le PGD pour la topologie de la convergence étroite sous des hypothèses d'existence, de continuité et de stricte positivité de la densité de transition, ainsi que de la tension exponentielle. La fonction de taux est l'entropie de Donsker-Varadhan au niveau-2:

$$J(\nu) := \sup \left\{ \int \log \frac{u}{Pu} d\nu;\ 1 \le u \in b\mathcal{B}(E) \right\},\ \forall \nu \in M_1(E) \tag{0.7}$$

où $b\mathcal{B}(E)$ est l'espace des fonctions réelles bornées et mesurables par rapport à la tribu de Borel $\mathcal{B}(E)$ de E, et $M_1(E)$ est l'espace de toutes les mesures de probabilité sur E. Par la suite, Deuschel et Stroock [27] ont établi le PGD pour la τ-topologie en supposant l'existence d'une dominée (que nous allons présenter plus bas).

Le PGD de l'estimateur à noyau de la densité pour la norme de L^1 est beaucoup plus fort et beaucoup plus difficile que le PGD des mesures d'occupation pour la τ-topologie. Cela nous pousse à chercher la condition de dépendance convenable et de nouveaux outils. Les résultats du chapitre 3 sont nouveaux . Quant aux points techniques, en comparaison du cas i.i.d., beaucoup plus d'outils dans l'analyse fonctionnelle et l'analyse convexe sont utilisés : ainsi une inégalité du type Harnack, un opérateur uniformément intégrable, la théorie des perturbation pour les opérateurs linéaires, le théorème de Bishop-Phelps, etc, sont-ils employés.

0.2.3 Méthode de partition de Devroye

Cette méthode a été introduite par Devroye [28] pour prouver la consistance de f_n^* dans le cas i.i.d.. Cette approche est réalisée par decomposition de l'espace $\mathbb{R}^d$ en petits rectangles ayant une maille qui dépendant de (h_n). Cette procédure permet de réduire l'estimation globale à l'estimation locale sur les petits rectangles. Tous les résultats connus sur les grandes déviations ou les déviations modérées pour f_n^* dépendent fortement de cette approche, ainsi que les nôtres dans cette thèse. Des précisions, nous invitons le lecteur à consulter les articles correspondants.

0.3 Présentation de principaux résultats

Nous présentons maintenant les résultats principaux de ce travail.

0.3.1 Cas i.i.d. (Chapitre 1)

Soit $\{X_i; i \geq 1\}$ une suite de variables aléatoires indépendantes et identiquement distribuées. Notons $\|\cdot\|_1 := \|\cdot\|_{L^1(\mathbb{R}^d,\, dx)}$. Nous considérons les deux questions suivantes:

Question 1. Quelle est l'estimation de grandes déviations de $\mathbb{P}\left(\|f_n^* - g\|_1 \leq \delta\right)$ pour une densité g fixée quelle conque?

Question 2. Pour chaque n fixé et chaque déviation $r > 0$, comment peut-on borner $\mathbb{P}\left(\|f_n^* - f\|_1 > r\right)$?

La Question 1 est importante pour les tests d'hypothèse et l'importance de la Question 2 est évidente à la fois en pratique et en théorie.

Le premier résultat presenté est théorique, puisqu'il étend le classique théorème de Sanov (voir Dembo et Zeitouni [25], Wu [93]).

Proposition 0.1 *Supposons que l'hypothèse* **(H1)** *est vérifiée et* $h_n \to 0$ *(sans* **(H2)***). Alors, lorsque n tend vers l'infini,* $\mathbb{P}(f_n^* \in \cdot)$ *satisfait un principe de grandes déviations (PGD) sur* $L^1(dx)$*, pour la topologie faible* $\sigma(L^1, L^\infty)$*), avec une fonction de taux donnée par :*

$$I(g) = \begin{cases} Ent_\mu(g) := \int_{\mathbb{R}^d} g(x) \log \frac{g(x)}{f(x)} dx, & si \;\; g \in \mathcal{P}, g(x)dx \ll f(x)dx; \\ +\infty & sinon \end{cases} \tag{0.8}$$

où $\mathcal{P}$ *est l'ensemble de toutes les fonctions de densité de probabilité sur* $\mathbb{R}^d$.

Remarques:

(a) Comme I n'est pas une bonne fonction de taux sur $L^1(\mathbb{R}^d)$ pour la topologie de la norme $\|\cdot\|_1$, $\mathbb{P}(f_n^* \in \cdot)$ ne satisfait pas le PGD sur $(L^1(\mathbb{R}^d), \|\cdot\|_1)$.

(b) Dans le cas i.i.d., on peut calculer la fonctionnelle de Cramér, donc utiliser le théorème de Gärtner-Ellis.

(c) Ce résultat ne donne aucune information pour les questions 1 and 2, parce que $\{\tilde{g} \in L^1(\mathbb{R}^d); \|\tilde{g} - g\|_1 < \delta\}$ n'est pas ouvert dans $(L^1(\mathbb{R}^d), \sigma(L^1, L^\infty))$, et $\{\tilde{g}; \|\tilde{g} - f\|_1 \geq r\}$ n'est pas fermé dans $(L^1(\mathbb{R}^d), \sigma(L^1, L^\infty))$.

Le résultat suivant dit quand même que $\mathbb{P}(f_n^* \in \cdot)$ satisfait le w*-PGD sur $(L^1(\mathbb{R}^d), \|\cdot\|_1)$. Cela répond à la Question 1).

Théorème 0.21 *Supposons que* **(H1)** *et* **(H2)** *sont vérifiées. Alors, pour tout* $g \in L^1(\mathbb{R}^d)$,

$$\begin{aligned} &\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\|f_n^* - g\|_{L^1(\mathbb{R}^d)} < \delta) \\ &= \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\|f_n^* - g\|_{L^1(\mathbb{R}^d)} < \delta) = -I(g). \end{aligned} \tag{0.9}$$

De plus pour tout sous-ensemble G **convexe** *dans* $L^1(\mathbb{R}^d)$ *et ouvert pour* $\|\cdot\|_1$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(f_n^* \in G) = -\inf_{g\in G} I(g). \tag{0.10}$$

Remarques:

(a) Par (0.9), pour tout sous-ensemble G dans $L^1(\mathbb{R}^d)$ ouvert pour $\|\cdot\|_1$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(f_n^* \in G) \geq -\inf_{g\in G} I(g).$$

En particulier, on a pour tout $r > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\|f_n^* - f\|_1 > r) \geq -\inf\{I(g);\ \|g - f\|_1 > r\}.$$

On peut montrer que $\inf\{I(g);\ \|g-f\|_1 > r\} = l(r)$ est exactement la fonction de taux trouvée par Louani [68]. Autrement dit, cette borne inférieure est beaucoup plus générale.

(b) Notons que $\{f_n^*, \|f_n^* - g\|_{L^1(\mathbb{R}^d)} < \delta\}$ était fermé pour la topologie $\sigma(L^1, L^\infty)$, on a alors la bornée supérieure dans ce théorème grâce à la Proposition 0.1. Pour la partie LLD, on est obligé d'adopter la méthode classique de “changement de mesure ”, car le théorème de Gärtner-Ellis ne s'applique pas ici.

Présentons maintenant la réponse à la Question 2.

Théorème 0.22 *Supposons que* **(H1)** *est vérifiée. Alors pour tous* $n \geq 1$ *et* $r > 0$,

$$\mathbb{P}(|D_n^* - \mathbb{E}D_n^*| > r) \leq 2\exp\left(-\frac{nr^2}{8}\right). \tag{0.11}$$

Comme $\mathbb{E}D_n^* \to 0$ sous les conditions **(H1)** et **(H2)**, l'inégalité ci-dessus est beaucoup plus précise que celle de [28]. En effect, pour chaque $r > 0$, il existe $C, \delta > 0$ tels que

$$\mathbb{P}\left(\|f_n^* - f\|_1 > r\right) \leq Ce^{-\delta n},\ \forall n \geq 1. \tag{0.12}$$

Dans Devroye [29], sous la condition **(H1)**, l'inégalité de déviation

$$\mathbb{P}(|D_n^* - \mathbb{E}D_n^*| > \epsilon) \le 2\exp(-\frac{n\epsilon^2}{32 \times 32})$$

est valable seulement dans le cas où ε est petit. Sa méthode repose sur la Poissonisation pour la queue d'échantillon de n. Notre méthode est complètement différente et repose sur l'inégalité de transport T1 du produit de mesures pour la distance de Hamming (voir l'excellente monographie de Ledoux [62]).

0.3.2 Cas des processus ϕ-mélangeants (Chapitre 2)

Soit $\{X_i, i \ge 1\}$ une suite retirée d'un processus $\phi-$mélangeant. Rappelons brièvement ce que veut dire la terminologie "processus ϕ-mélangeant". Soit une suite de variables aléatoires $(X_i)_{i\ge 1}$ à valeurs dans $\mathbb{R}^d$, définie sur $(\Omega, \mathcal{F}, \mathbb{P})$. Pour deux sous-tribus $\mathcal{A}, \mathcal{B}$ de $\mathcal{F}$, définissons

$$\phi(\mathcal{A}, \mathcal{B}) = \sup\left\{\left|\mathbb{P}(V) - \frac{\mathbb{P}(U \bigcap V)}{\mathbb{P}(U)}\right|;\ U \in \mathcal{A}, \mathbb{P}(U) \ne 0, V \in \mathcal{B}\right\}.$$

Et pour chaque entier k positif,

$$\phi_k := \sup_{m\ge 1}\left\{\phi(\sigma(X_1, \cdots, X_m), \sigma(X_{m+l}; l \ge k))\right\}.$$

On dit que (X_n) est ϕ-mélangeante , si $\lim_{k\to+\infty}\phi_k = 0$.

Rappelons que même pour les chaînes de Markov stationnaires et récurrentes de Doeblin (cas dans lequel, ϕ_k décroit exponentiellement vite vers zéro), le PGD pour les mesures empiriques n'est pas vrai en général (voir Bryc et Dembo [11]). Donc on n'espère pas qu'il soit vrai pour f_n^* dans $L^1(\mathbb{R}^d)$ dans le cas ϕ-mélangeant, à moins d'imposer la condition de décroissance exponentielle supérieurement de ϕ_k comme dans l'article de Bryc et Dembo [11].

On faisons un premier pas pour les grandes déviations de f_n^* : en établissant la convergence exponentielle de f_n^* vers f dans $L^1(\mathbb{R}^d, dx)$ pour les suites ϕ-mélangeantes satisfaisant $\sum_k \phi_k < +\infty$, sous les conditions **(H1)** et **(H2)**. De plus, on obtient une inégalité exponentielle du type de Hoeffding.

Nos résultats principaux sont:

Théorème 0.23 *Soit $(X_i)_{i\in\mathbb{N}^*}$ une suite stationnaire de variables aléatoires à valeurs dans $\mathbb{R}^d$ possédant une loi marginale $\mu(dx) = f(x)dx$. Supposons que*

$$S_\phi := \sum_{k=1}^{\infty}\phi_k < +\infty. \tag{0.13}$$

*Soit K une fonction mesurable positive sur $\mathbb{R}^d$ telle que $\int K(x)dx = 1$ (**(H1)**) et (h_n) une suite de nombres strictement positifs vérifiant **(H2)**. Alors $D_n^* \to 0$ exponentiellement vite lorsque $n \to \infty$, c'est-à-dire,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(D_n^* > \delta) < 0, \qquad \forall \delta > 0.$$

Théorème 0.24 *Dans le contexte du Théorème 0.23, supposons que (0.13) et* **(H1)** *sont vérifiées. Alors, pour tout $n \geq 1$ et tout $r > 0$,*

$$\mathbb{P}(|D_n^* - \mathbb{E}D_n^*| > r/\sqrt{n}) \leq 2\exp\left(-\frac{r^2}{2(1+2S_\phi)^2}\right). \tag{0.14}$$

Tous ces résultats reposent fortement sur l'inégalité suivante du type de Hoeffding, établie récemment par Rio ([81]).

Lemme 0.25 *([81]) Soit $f : E^n \to \mathbb{R}$ satisfaisant*

$$|f(x) - f(y)| \leq L \tag{0.15}$$

pour tous $x, y \in E^n$ vérifiant $\#\{i;\ x_i \neq y_i\} = 1$. Alors,
$\forall \lambda > 0$,

$$\begin{aligned} &\mathbb{E}\exp\left[\lambda\left(f(X_1,\cdots,X_n) - \mathbb{E}f(X_1,\cdots,X_n)\right)\right] \\ &\quad \leq \exp\left(\frac{\lambda^2}{8}\cdot nL^2(1+2S_\phi)^2\right); \end{aligned} \tag{0.16}$$

et en particulier $\forall t > 0$,

$$\mathbb{P}\left(f(X_1,\cdots,X_n) - \mathbb{E}f(X_1,\cdots,X_n) > t\right) \leq \exp\left(-\frac{2t^2}{nL^2(1+2S_\phi)^2}\right). \tag{0.17}$$

Considérons la distance de Hamming sur E^n:

$$d_H(x,y) := \#\{i;\ x_i \neq y_i\}.$$

La condition (0.15) est équivalente à

$$|f(x) - f(y)| \leq L d_H(x,y),\ \forall x, y \in E^n$$

c'est-à-dire, le coefficient lipschitzien de f pour la distance de Hamming d_H est plus petit que L. On peut donc alors traduire l'inégalité de Rio (0.16) par l'inégalité de transport suivante (Bobkov-Götze [8]) :

Corollaire 0.26 *Soit μ_n la loi de $(X_1, \cdots, X_n)$. Alors pour toute mesure de probabilité ν sur E^n,*

$$W_1(\nu, \mu_n) \leq (1 + 2S_\phi)\sqrt{\frac{n}{2} \cdot h(\nu; \mu_n)}. \tag{0.18}$$

où $W_1(\nu; \mu_n)$ est la distance de Wasserstein entre ν et μ_n, définie par

$$W_1(\nu; \mu_n) := \inf \iint d_H(x, y) d\pi(x, y),$$

l'infimum étant pris sur toutes les mesures de probabilité π sur $E^n \times E^n$ possédant les lois marginales ν et μ_n; et

$$h(\nu, \mu_n) := \begin{cases} \int \log \frac{d\nu}{d\mu_n} d\nu & si\ \nu \ll \mu_n, \\ +\infty & sinon \end{cases}$$

est l'entropie relative (ou l'information de Kullback) de ν par rapport à μ_n.

Remarquons que si $X_1, \cdots, X_n$ sont indépendantes, alors $S_\phi = 0$ et l'inégalité (0.18) est une conséquence de l'inégalité de Pinsker ainsi que de technique des tenseurs. Cette inégalité (0.18) a d'abord été obtenue par Marton [69] pour des chaînes de Markov récurrentes de Doeblin, et puis elle a été étendue par Samson [82] pour des suites générales de variables aléatoires ϕ-mélangeantes. Mais la condition apparaisant dans ce travail [82] est $\sum_k \sqrt{\phi_k} < +\infty$, et elle est plus forte que celle-ci.

Le lecteur est renvoyé à Ledoux [62] pour un traitement systématique (et aussi des références abondantes) et pour des applications d'une telle inégalité de transport aux mesures de concentration, ainsi qu'à Djellout, Guillin et Wu [30] pour les développements récents des inégalités de transport.

0.3.3 Cas des processus de Markov uniformément ergodiques (Chapitre 3)

Soit $\{X_n; n \geq 0\}$ une chaîne de Markov récurrente de Doeblin à valeurs dans un sous-ensemble borélien E mesurable de $\mathbb{R}^d$, définie sur un espace de probabilité $(\Omega, (\mathcal{F}_n)_{n\in\mathbb{N}}, \mathcal{F}, (\mathbb{P}_x)_{x\in E})$, et possédant une densité de transition $P(x, dy)$ (inconnue). Nous supposons de plus que la seule mesure invariante μ de P est absolument continue, i.e., $\mu(dx) = f(x)dx$ la densité f est inconnue.

Dans cette partie, on utilisons les notations suivantes : $L^p(\mathbb{R}^d) := L^p(\mathbb{R}^d, dx)$, $L^p(\mu) := L^p(E, \mu)$; $\|f\|_1 = \|f\|_{L^1(\mathbb{R}^d, dx)}$. On note $b\mathcal{B}$ (resp. $b\mathcal{B}(E)$) l'espace de toutes les fonctions réelles bornées et boréliennes $\mathcal{B}$-mesurables sur $\mathbb{R}^d$ (resp. E, muni de la norme $\|V\| = \sup_x |V(x)|$. Et $M_1(E)$ est l'espace de toutes les mesures de probabilité sur E. On écrit $\nu(V) = \langle V \rangle_\nu := \int_E V(x)d\nu(x)$. Sans perdre la généralité, nous supposons que $(X_n)_{n\geq 0}$ est le système de coordonnées sur $\Omega := E^{\mathbb{N}}$ et $\mathbb{P}_x$ est la loi d'une chaîne de Markov qui a la densité de transition P et un point initial $x \in E$. Nous définissons $\mathbb{P}_\nu(\cdot) := \int_E \mathbb{P}_x(\cdot)d\nu(x)$ et $\mathbb{E}^\nu(\cdot) = \int_\Omega \cdot\, d\mathbb{P}_\nu$. Enfin $(\theta\omega)_n := \omega_{n+1}(n \in \mathbb{N})$ est l'opérateur de shift sur Ω.

Les grandes déviations des mesures d'occupation $L_n := (1/n)\sum_{k=0}^{n-1} \delta_{X_k}$ pour des processus de Markov sont un sujet traditionnel en probabilités, qui a été étudié d'abord par Donsker et Varadhan [32]. La fonction de taux est l'entropie de Donsker-Varadhan au niveau-2 ci-dessous:

$$J(\nu) := \sup\left\{\int \log\frac{u}{Pu}d\nu;\ 1 \leq u \in b\mathcal{B}(E)\right\},\ \forall \nu \in M_1(E) \tag{0.19}$$

Deuschel et Stroock [27] (Thm 4.1.14) ont obtenu le PGD de L_n par rapport à la τ-topologie (c'est-à-dire, la topologie la plus faible sur $M_1(E)$ telle que $\nu \to \nu(f) := \int_E f(x)d\nu(x)$ est continue pour toute fonction $f \in b\mathcal{B}(E)$), sous l'hypothèse suivante :

(H) (*uniforme ergodicité*) Il existe $1 \leq l \leq N \in \mathbb{N}$ et $M \geq 1$ tels que

$$P^l(x, A) \leq M\frac{P(y, A) + \cdots + P^N(y, A)}{N},\ \forall x, y \in E, A \in \mathcal{B}(E).$$

Beaucoup de progrès notables ont été réalisés ensuite, voir par exemple [25], [94], [73] et les références qu'ils contiennent.

Nos résultats principaux sont les suivants:

Théorème 0.27 *Supposons que* **(H)** *et* $h_n \to 0$ *(sans* **(H2)***) sont vérifiées. Alors* $\mathbb{P}_x(f_n^* \in \cdot)$ *satisfait, uniformément pour les points initiaux* $x \in E$*, le PGD dans* $L^1(\mathbb{R}^d)$ *pour la topologie faible* $\sigma(L^1, L^\infty)$*, avec fonction de taux ci-dessous :*

$$J(g) := \begin{cases} J(gdx), & si\ \ g \in \mathcal{P}(E) \\ +\infty, & si\ \ g \in L^1(\mathbb{R}^d)\backslash\mathcal{P}(E). \end{cases} \tag{0.20}$$

Ici $J(\cdot)$ *est l'entropie de Donsker-Varadhan au niveau-2 explicité dans (0.19),* $\mathcal{P}(E)$ *est l'ensemble de toutes les densités de probabilité sur* $\mathbb{R}^d$ *ayant leurs supports dans* E*, c'est-à-dire, des fonctions qui ont les propriétés suivantes :*
$g \in L^1(\mathbb{R}^d)$ *telle que* $g \geq 0$ *sur* $\mathbb{R}^d$*,* $g = 0$ *p.s. sur* $E^c := \mathbb{R}^d\backslash E$*, et* $\int_{\mathbb{R}^d} gdx = 1$.

Plus précisément , J est inf-compacte sur $(L^1(\mathbb{R}^d), \sigma(L^1, L^\infty))$, et pour tout ensemble mesurable A de $L^1(\mathbb{R}^d)$,

$$\begin{aligned} -\inf_{g\in \overset{\circ}{A}^\sigma} J(g) &\le \liminf_{n\to\infty} \frac{1}{n} \log \inf_{x\in E} \mathbb{P}_x(f_n^* \in A) \\ &\le \limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in E} \mathbb{P}_x(f_n^* \in A) \le -\inf_{g\in \bar{A}^\sigma} J(g) \end{aligned}$$

où $\overset{\circ}{A}^\sigma, \bar{A}^\sigma$ correspondent respectivement à l'intérieur et la fermeture de A pour la topologie faible $\sigma(L^1, L^\infty)$.

Remarques:

(a) Le Théorème 0.27 ressemble beaucoup au résultat classique de Deuschel et Stroock [27] de L_n pour la τ-topologie. La preuve est basé sur une inégalité du type Harnack pour le rapport entre les fonctions propres de l'opérateur de Feynman-Kac P^V et le rayon spectral de cet opérateur $e^{\Lambda(V)}$, ainsi que sur la continuité *monotone* de la fonctionnelle de Cramér $\Lambda(V)$. La continuité de $\Lambda(V)$ est vérifiée grâce aux résultats sur “ les opérateurs uniformément intégrables ” développés par Wu [94].

(b) Le PGD pour la topologie faible sur $L^1(\mathbb{R}^d)$, ci-dessus dessus est très faible au sens où on n'a pas la consistance (D_n^* tend vers 0 en probabilités). Pour les applications en statistique, les quantités principales à étudier sont

 (i) $\mathbb{P}_x(\|f_n^* - g\|_1 < \delta)$ $g \in \mathcal{P}(E)$ étant fixée. Cet étude est importante pour les tests d'hypothèses : (H_0) : $d\mu(x) = f(x)dx$ contre (H_1) : $d\mu(x) = g(x)dx$;

 (ii) $\mathbb{P}_x(D_n^* > \delta)$, qui sert en statistique.

 Malheureusement, le Théorème 0.27 ne peut pas servir à cette étude, parce que l'ensemble $\{\tilde{g} \in L^1(\mathbb{R}^d);\ \|\tilde{g} - g\|_1 < \delta\}$ n'est pas ouvert dans $\sigma(L^1, L^\infty)$ et que l'ensemble $\{\tilde{g} \in L^1(\mathbb{R}^d);\ \|\tilde{g} - f\|_1 \ge \delta\}$ n'est pas fermé dans $\sigma(L^1, L^\infty)$.

Théorème 0.28 *Supposons que les hypothèses* **(H)** *et* **(H2)** *sont vérifiées. Alors, $\mathbb{P}_x(f_n^* \in \cdot)$ satisfait, uniformément pour tous les états initiaux $x \in E$, le w*-PGD sur $(L^1(\mathbb{R}^d), \|\cdot\|_1)$ de fonction de taux $J(g)$ donnée par (0.20). C'est-à-dire que pour tout $g \in L^1(\mathbb{R}^d)$,*

$$\begin{aligned} &\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \inf_{x\in E} \mathbb{P}_x(\|f_n^* - g\|_1 < \delta) \\ &= \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in E} \mathbb{P}_x(\|f_n^* - g\|_1 < \delta) = -J(g). \end{aligned} \tag{0.21}$$

Remarques:

(a) La borne supérieure dans le Théorème 0.28 est une conséquence du Théorème 0.27. Mais pour la borne inféreure dans le Théorème 0.28, puisque f_n^* n'est pas (en général) tendue exponentiellement dans $(L^1, \|\cdot\|_1)$ (sinon, on a inf-compactité de la fonction de taux J pour la norme de $\|\cdot\|_1$, qui est faux dans le contexte présenté), donc le théorème de Gärtner-Ellis abstrait due à Baldi [25] ne s'applique pas ici. Dans ce cas là, pour la borne inférieure, on est obligé d'adopter la méthode classique du"changement de mesure", les résultats principaux de notre acticle précédent [64] (Théorème 0.23) et le théorème de Bishop-Phelps.

(b) Le bon PGD correspondant est faux en général, parce que même dans le cas i.i.d., la fonction $J(g) = J^{iid}(g) = \int g(x) \log \frac{g(x)}{f(x)} dx$ (où $g \in \mathcal{P}(E)$ et $gdx \ll fdx$) n'est pas inf-compact sur $(L^1(\mathbb{R}^d), \|\cdot\|_1)$ (comme noté dans [63]).

Théorème 0.29 *Supposons que les hypothèses* **(H)** *et* **(H2)** *sont vérifiées. Alors,*

(a) Pour tout $\delta > 0$,

$$\begin{aligned} -I(\delta) &\le \liminf_{n\to\infty} \frac{1}{n} \log \inf_{x\in E} \mathbb{P}_x(\|f_n^* - f\|_1 > \delta) \\ &\le \limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in E} \mathbb{P}_x(\|f_n^* - f\|_1 > \delta) \le -I(\delta-) \end{aligned} \tag{0.22}$$

où

$$I(\delta) = \inf\{J(g) | g \in \mathcal{P}(E), \|g - f\|_1 > \delta\}, \tag{0.23}$$

($I(\delta-)$ étant la limite à gauche en δ de I).

(b) Pour tout $\delta > 0$,

$$I(\delta) \ge \frac{1}{l}\left(I^{iid}(\delta) - \log M\right) \tag{0.24}$$

où l, M sont donnés par **(H)**, *et $I^{iid}(\delta)$ est la fonction de taux du PGD pour $\|f_n^* - f\|_1$ dans le cas où la suite (X_n) est i.i.d. de loi commune μ.*

(c) Outre l'hypothèse **(H)**, *supposons que P est apériodique. Alors, on a de plus*

$$I(\delta) \ge \frac{\delta^2}{2(1+S)^2}, \ \forall \delta > 0 \tag{0.25}$$

où $S := \sum_{k=1}^{\infty} \sup_{x,y\in E} \|P^k(x,\cdot) - P^k(y,\cdot)\|_{TV}$ (ici $\|\cdot\|_{TV}$ dénote la variation totale) est fini.

Remarques:
La borne inférieure dans le Théorème 0.29 est une conséquence facile du Théorème 0.27. La borne supérieure dans le Théorème 0.29 est beaucoup plus difficile. Elle est basé sur une extension de l'inégalité de Cramér, qui est une des inégalités de déviations les plus précises dans le cas i.i.d., (voir Lemme 3.2 dans Chapitre 3). Grâce à cette inégalité de déviations, le Théorème 0.28 peut être vérifié directement par la méthode de partition de Devroye [28] et Louani [68].

Grâce aux résultats ci-dessus, on a déjà l'estimation de grandes déviations pour l'estimateur f_n^*, ce qui est pratique en statistiques. Nous monterons maintenant que f_n^* est optimal asymptotiquement au sens de Bahadur. Soit Θ un ensemble de données inconnues, (P,μ) satisfaisant **(H)** et $\mu(dx) \ll dx$. Etant donné un sous-ensemble $\mathcal{D}$ de la boule unité dans $b\mathcal{B}$, on dit qu'un estimateur $T_n(\cdot) := T_n(\cdot; X_0, \cdots, X_{n-1}) \in L^1(\mathbb{R}^d)$ est un estimateur asymptotiquement $\sigma(L^1, \mathcal{D})$-consistant de la densité f, si $\forall V \in \mathcal{D}$,

$$\int_{\mathbb{R}^d} T_n(x)V(x)dx \to \int_{\mathbb{R}^d} f(x)V(x)dx$$

en probabilité $\mathbb{P}_\mu$. On a alors : :

Théorème 0.30 *Etant donné $(P,\mu) \in \Theta$, soit $((X_n), (\mathbb{P}_x)_{x\in E})$ le processus de Markov associé.*

(a) **(Borne inférieure de type Bahadur)** *Supposons que $\mathcal{D}$ est dense dans la boule unité de $L^\infty(\mathbb{R}^d)$ pour la topologie faible* $\sigma(L^\infty, L^1)$. Alors pour tout estimateur asymptotiquement $\sigma(L^1, \mathcal{D})$-consistant de la densité f,*

$$\begin{aligned} &\liminf_{r\to 0+} \frac{1}{r^2} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\|T_n - f\|_1 > r) \\ &\geq -\frac{1}{2\sup_{\|V\|\leq 1} \sigma^2(V)} = -\frac{1}{8\sup_{A\in\mathcal{B}} \sigma^2(1_A)} \end{aligned} \tag{0.26}$$

où

$$\sigma^2(V) := Var_\mu(V) + 2\sum_{k=1}^\infty \langle V - \mu(V), P^k V\rangle_\mu.$$

De plus, si $\|T_n - T_n \circ \theta^N\|_1 \leq \delta_n \to 0$, alors (0.26) est encore vrai avec $\inf_{x\in E} \mathbb{P}_x$ à la place de $\mathbb{P}_\mu$.

(b) **(Efficacité asymptotique de f_n^* au sens de Bahadur)** *Si h_n vérifie* **(H2)**,

alors

$$\begin{aligned}
&\liminf_{r\to 0+}\frac{1}{r^2}\lim_{n\to\infty}\frac{1}{n}\log\inf_{x\in E}\mathbb{P}_x(\|f_n^*-f\|_1>r)\\
&=\limsup_{r\to 0+}\frac{1}{r^2}\lim_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{P}_x(\|f_n^*-f\|_1>r)\\
&=-\frac{1}{2\sup_{\|V\|\le 1}\sigma^2(V)}=-\frac{1}{8\sup_{A\in\mathcal{B}}\sigma^2(1_A)}.
\end{aligned}\tag{0.27}$$

Remarques:

(a) Par conséquent f_n^* est un estimateur asymptotiquement efficace de f au sens de Bahadur. Ainsi $1/\sigma^2(V)$ peut être vu comme l'information de Fisher dans la direction V de notre modèle statistique Θ.

(b) L'optimalité asymptotique de f_n^* au sens de Bahadur montrée dans le Théorème 0.30 et l'estimation précise du risque minimal (0.27) sont basées sur la théorie analytique des perturbations des opérateurs, due à Kato [56].

(c) Tous les résultats ci-dessus, pour autant que nous le sachions, sont nouveaux dans le cas dépendant.

0.3.4 Cas des processus de Markov réversibles (Chapitre 4)

L'ergodicité uniforme **(H)** n'est pas satisfaite en général pour des modèles concrets avec des espaces d'état non-compacts. Par exemple, tous les processus de Markov Gaussiens, stationnaires et ergodiques à valeurs dans $\mathbb{R}$ sont réversibles mais pas uniformément ergodiques. Dans cette section, nous nous plaçons dans le cas d'un processus de Markov réversible, possédant un noyau de transition uniformément intégrable dans L^2. Nous rappelons, pour commencer, la notion d'opérateurs uniformément intégrables, proposée par Wu [94].

Définition 0.31 *(a) Soit $p\in[1,+\infty)$. Un opérateur borné linéaire $\pi: L^p(\mu)\to L^p(\mu)$ est dit* ***uniformément intégrable*** *dans $L^p(\mu)$, si $\pi(B^p(1))$ est uniformément intégrable dans $L^p(\mu)$, où $B^p(1)=\{f\in L^p(\mu)\mid \|f\|_{L^p(\mu)}\le 1\}$.*

(b) Soit $p=+\infty$. Un opérateur borné linéaire $\pi: L^\infty(\mu)\to L^\infty(\mu)$ est dit ***uniformément intégrable*** *dans $L^\infty(\mu)$, si pour chaque suite $(A_n)\subset\mathcal{A}$ tendant vers $\varnothing$,*

$$\{\|\pi(1_{A_n})\|\}_\infty\overset{n\to+\infty}{\longrightarrow}0.$$

Pour un exposé des propriétés des opérateurs uniformément ergodiques , voir Wu [94]. Lorsque le noyau de transition P possède une mesure μ de probabilité invariante, l'ergodicité uniforme **(H)** est beaucoup plus forte que l'intégrabilité uniforme de P^l dans $L^p(\mu)$ pour chaque $1 < p < +\infty$. Un exemple typique est un processus Gaussien stationnaire et ergodique à valeurs dans $\mathbb{R}^d$: son noyau de transition P est toujours uniformément intégrable dans $L^2(\mu)$, mais pas uniformément ergodique au sens de **(H)**.

Soit $\{X_n; n \geq 0\}$ une chaîne de Markov réversible à valeurs dans $\mathbb{R}^d$, définie sur l'espace de probabilité $(\Omega, (\mathcal{F}_n^0)_{(n\in\mathbb{N})}, \mathcal{F}, (\mathbb{P}_x)_{x\in\mathbb{R}^d})$, et possédant un noyau de transition de Markov $P(x, dy)$ (inconnu). Supposons que :

(A1) P est irréductible (Meyn et Tweedie [71]) et symétrique par rapport à l'unique mesure de probabilité invariante μ, qui est absolument continue , c'est-à-dire, $d\mu(x) = f(x)dx$, où la densité f est inconnue ;

(A2) Pour un certain $N \geq 1$, P^N est uniformément intégrable dans $L^2(\mu)$, c'est-à-dire, $\{(P^N f)^2;\ \|f\|_{L^2(\mu)} \leq 1\}$ est uniformément intégrable.

Remarque:
Wu [94] a montré que **(A2)** est une condition suffisante pour obtenir le PGD de L_n dans l'espace $M_1(\mathbb{R}^d)$ des mesures de probabilité sur $\mathbb{R}^d$ pour la τ-topologie, et que cette condition était même nécessaire dans le cas réversible (voir Wu[97]). La fonction de taux est alors donnée par

$$J_\mu(\nu) := \begin{cases} \sup\left\{\int \log \frac{u}{Pu} d\nu;\ 1 \leq u \in b\mathcal{B}\right\}, & \forall \nu \in M_1(\mathbb{R}^d), \nu \ll \mu; \\ +\infty, & \text{sinon.} \end{cases} \tag{0.28}$$

Etant donné un échantillon observé $\{X_0, \ldots, X_n\}$, considérons la mesure empirique du type trapèze, c'est-à-dire,

$$L_n = \frac{1}{n}\left(\sum_{i=1}^{n-1} \delta_{X_i} + \frac{1}{2}(\delta_{X_0} + \delta_{X_n})\right) = \frac{1}{n}\sum_{i=0}^{n-1} \frac{1}{2}\left(\delta_{X_i} + \delta_{X_{i+1}}\right),$$

Soit $K : \mathbb{R}^d \to \mathbb{R}$ une fonction mesurable telle que

$$K \geq 0,\ \int_{\mathbb{R}^d} K(x)dx = 1, \tag{0.29}$$

et posons $K_h(x) = \dfrac{1}{h^d} K(\dfrac{x}{h})$ pour tout $h > 0$. L'estimateur à noyau de la densité de la fonction inconnue f est défini ci-dessous. Pour tout $x \in \mathbb{R}^d$,

$$f_n^*(x) := K_{h_n} * dL_n(x) = \frac{1}{n}\sum_{i=0}^{n-1} \frac{1}{2h_n^d}\left(K\left(\frac{x - X_i}{h_n}\right) + K\left(\frac{x - X_{i+1}}{h_n}\right)\right), \tag{0.30}$$

où $h = h_n, \{h_n, n \geq 0\}$ est une suite de nombres strictement positifs (la fenêtre) satisfaisant

$$h_n \to 0, \quad nh_n^d \to +\infty \quad lorsque \quad n \to \infty. \tag{0.31}$$

Dans cette section, on adoptons la notation suivante :

$$L^p := L^p(\mathbb{R}^d) := L^p(\mathbb{R}^d, dx), \ \|f\|_p = \|f\|_{L^p(\mathbb{R}^d,dx)}, \ L^p(\mu) := L^p(\mathbb{R}^d, \mu).$$

Et pour tout $L \geq 1$, on note :

$$\mathcal{A}_{\mu,2}(L) := \left\{\nu \in M_1(\mathbb{R}^d); \ \nu \ll \mu, \|\frac{d\nu}{d\mu}\|_{L^2(\mu)} \leq L\right\}, \ \mathcal{A}_{\mu,2} := \bigcup_{L\geq 1} \mathcal{A}_{\mu,2}(L).$$

Théorème 0.32 *Supposons que les hypothèses* **(A1)** *et* **(A2)** *sont vérifiées, et que $h_n \to 0$ (sans (0.31)). Alors pour tout $L \geq 1$, $\mathbb{P}_\nu(f_n^* \in \cdot)$ satisfait, uniformément pour toutes les mesures initiales $\nu \in \mathcal{A}_{\mu,2}(L)$, le PGD dans L^1 pour la topologie faible $\sigma(L^1, L^\infty)$ de fonction de taux*

$$J(g) := \begin{cases} J(gdx), & si \ \ gdx \in M_1(\mathbb{R}^d) \ \ et \ \ gdx \ll fdx; \\ +\infty, & sinon, \end{cases} \tag{0.32}$$

où $J(\cdot)$ étant l'entropie de Donsker-Varadhan donnée dans (0.28). Plus précisément, J est inf-compacte sur $(L^1, \sigma(L^1, L^\infty))$, et pour tout sous-ensemble mesurable A dans L^1, pour tout $L \geq 1$,

$$\begin{aligned} -\inf_{g\in \overset{\circ}{A}^\sigma} J(g) &\leq \liminf_{n\to\infty} \frac{1}{n} \log \inf_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(f_n^* \in A) \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(f_n^* \in A) \leq -\inf_{g\in\bar{A}^\sigma} J(g) \end{aligned}$$

où $\overset{\circ}{A}^\sigma, \bar{A}^\sigma$ correspondent à l'intérieur et la fermeture de A pour la topologie faible $\sigma(L^1, L^\infty)$.

Remarques:

(a) Pour obtenir la borne supérieure, on applique le théorème généralisé de Gärtner-Ellis et les résultats (importants) établis pour un opérateur uniformément intégrable dans [94].

(b) Pour la borne inférieure , la méthode classique de changement de mesure est utilisée, ainsi qu'une approximation de mesure (le théorème de Bishop-Phelps) comme dans le cas uniformément ergodique.

(c) Le PGD pour la topologie faible sur L^1 ci-dessus est très faible qu'il ne montre pas $D_n^* \to 0$ en probabilités.

Nous passons maintenant au :

Théorème 0.33 *Supposons que* **(A1)** *et* **(A2)** *sont vérifiées, ainsi que (0.31). Alors, pour tous $L \geq 1$ et $\delta > 0$,*

$$\begin{aligned} -I(\delta) &\leq \liminf_{n\to\infty} \frac{1}{n} \log \inf_{\nu \in \mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - f\|_1 > \delta) \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu \in \mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - f\|_1 \geq \delta) \leq -I(\delta-) \end{aligned} \tag{0.33}$$

où

$$I(\delta) = \inf\{J(g) | g \in L^1, \|g - f\|_1 > \delta\} > 0 \tag{0.34}$$

et $I(\delta-)$ est la limite à gauche de I en δ.

Remarques: Ce théorème est basé sur une inégalité de déviations du type Cramér, voir Lemme 3.1 dans l'article. L'inégalité de déviations est fortement basée sur le fait que le rayon spectral pour un opérateur symétrique L^2 est égal à la norme de cet opérateur dans L^2. Au vu de cette nouvelle inégalité, nous pouvons maintenant utiliser de façon valable l'approche de partition de Devroye [28] et Louani [68].

Théorème 0.34 *Supposons que* **(A1)** *et* **(A2)** *sont vérifiées. Supposons aussi (0.31). Alors pour tout $L \geq 1$ $\mathbb{P}_\nu(f_n^* \in \cdot)$ satisfait le w^*-PGD de fonction de taux J sur $(L^1, \|\cdot\|_1)$ uniformément sur toutes les mesures initiales $\nu \in \mathcal{A}_{\mu,2}(L)$. C'est-à-dire, pour tout $L \geq 1$ et $g \in L^1$,*

$$\begin{aligned} &\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \inf_{\nu \in \mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - g\|_1 < \delta) \\ &\quad = \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu \in \mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - g\|_1 < \delta) = -J(g). \end{aligned} \tag{0.35}$$

Remarques:

(a) La borne supérieure dans le Théorème 0.34 est une conséquence du Théorème 0.32.

(b) Mais pour la borne inférieure dans le Théorème 0.34, on utilise la méthode du changement de mesure ainsi que le théorème de Bishop-Phelps.

Comme dans le cas uniformément ergodique, on montre que f_n^* est optimal asymptotiquement au sens de Bahadur.

Soit Θ un ensemble de données inconnues (P,μ) satisfaisant **(A1)** et **(A2)**. Etant donné un sous-ensemble $\mathcal{D}$ de la boule unité dans $b\mathcal{B}$, on dit qu'un estimateur $T_n(\cdot) := T_n(\cdot; X_0, \cdots, X_{n-1}) \in L^1(\mathbb{R}^d, dx)$ est un estimateur asymptotiquement $\sigma(L^1, \mathcal{D})$-consistant de la densité f, si pour tout $V \in \mathcal{D}$, $\int_{\mathbb{R}^d} T_n(x)V(x)dx \to \int_{\mathbb{R}^d} f(x)V(x)dx$ en mesure de probabilité $\mathbb{P}_\mu$.

Etant donné $(P,\mu) \in \Theta$, soit $((X_n), (\mathbb{P}_x)_{x\in E})$ le processus de Markov associé.

Théorème 0.35 *(a)* **(Borne inférieure du type Bahadur)** *Supposons que $\mathcal{D}$ est dense dans la boule unité de $L^\infty(\mathbb{R}^d)$ pour la topologie faible* $\sigma(L^\infty, L^1)$. Alors pour tout estimateur asymptotiquement $\sigma(L^1, \mathcal{D})$-consistant de la densité f ,*

$$\begin{aligned}
&\liminf_{r\to 0+} \frac{1}{r^2} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\|T_n - f\|_1 > r) \\
&\quad \geq -\frac{1}{2\sup_{\|V\|\leq 1} \sigma^2(V)} = -\frac{1}{8\sup_{A\in\mathcal{B}} \sigma^2(1_A)},
\end{aligned} \tag{0.36}$$

où

$$\sigma^2(V) := 2\sum_{k=0}^{\infty} \langle V, P^k(V - \mu(V))\rangle_\mu - Var_\mu(V).$$

De plus, si $\|T_n - T_n \circ \theta^N\|_1 \leq \delta_n \to 0$, alors (0.36) est encore vrai avec $\mathbb{P}_\nu$ à la place de $\mathbb{P}_\mu$ pour toutes les mesures initiales $\nu \in M_1(E)$, θ étant l'opérateur de shift sur Ω.

(b) **(Efficacité asymptotique de f_n^* au sens de Bahadur)** *Si h_n vérifie (0.31), alors*

$$\begin{aligned}
&\liminf_{r\to 0+} \frac{1}{r^2} \liminf_{n\to\infty} \frac{1}{n} \log \inf_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - f\|_1 > r) \\
&= \limsup_{r\to 0+} \frac{1}{r^2} \limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - f\|_1 > r) \\
&= -\frac{1}{2\sup_{\|V\|\leq 1} \sigma^2(V)} = -\frac{1}{8\sup_{A\in\mathcal{B}} \sigma^2(1_A)}.
\end{aligned} \tag{0.37}$$

Remarque:

Par conséquent, f_n^* est un estimateur asymptotiquement efficace de f au sens de Bahadur. Ainsi $1/\sigma^2(V)$ peut être vu comme l'information de Fisher dans la direction V de notre modèle statistique Θ.

0.3.5 Consistance forte et TCL pour l'estimateur de décrément aléatoire (Chapitre 5)

Pour la maintenance et la détection de dégâts de grandes infrastructures (les ponts suspendus, par exemple), il n'est pas possible d'utiliser les excitations contrôlées pour au moins deux raisons : les travaux trop difficiles à réaliser et l'impossibilité de mettre l'infrastructure hors-service pendant trop longtemps. Par conséquent, les techniques utilisant les réponses dynamiques sous les chargements ambiants (comme les vents et le traffic) présentent un grand intérêt. C'est bien le cas de l'algorithme de décrément aléatoire, introduit par Cole [19] dans les années 60. Une investigation rigoureuse des propriétés de l'estimateur de décrément aléatoire (abrégé en EDA dans la suite) devient une question très intéressante et importante.

On travaille sous le cadre suivant :
Soit $\{X_i, i \geq 0\}$ une suite gaussienne, stationnaire et ergodique de moyenne nulle et de variance σ^2 où $\sigma > 0$. Soit

$$\rho_j := \rho(j) = \frac{Cov(X_j, X_0)}{\sqrt{Var(X_j)Var(X_0)}}$$

le coefficient de corrélation entre X_j et X_0 pour tout entier j.

La condition suivante est connue sous le nom de " **condition d'atteinte**" dans la méthode de décrément aléatoire :

$$\mathcal{D}_k = (X_k, X_{k+1}, \ldots, X_{k+d-1}) \in \Delta; \tag{0.38}$$

où d est un entier fixé et Δ est un domaine dans $\mathbb{R}^d$. Pour les applications en mécanique, $d = 1$ ou $d = 2$.

Soient $(\tau_k, k \in \mathbb{Z})$ les moments successifs des passages dans le domaine Δ, plus précisément pour $k > 0$:

$$\begin{aligned} \tau_1 &= \inf\{j \geq 0 : (X_j, X_{j+1}, \ldots, X_{j+d-1}) \in \Delta\}, \\ \tau_{k+1} &= inf\{j > \tau_k : (X_j, X_{j+1}, \ldots, X_{j+d-1}) \in \Delta\}; \end{aligned} \tag{0.39}$$

L'estimateur de décrément aléatoire est défini par

$$D_n(j) := \frac{1}{n}\sum_{k=1}^{n} X_{\tau_k+j}, \tag{0.40}$$

ou par

$$\bar{D}_n(j) := \frac{1}{l(n)}\sum_{k=1}^{l(n)} X_{\tau_k+j}, \ l(n) = \sum_{k=0}^{n-1} 1_\Delta(X_j, X_{j+1}, \ldots, X_{j+d-1}). \tag{0.41}$$

Introduisons maintenant notre condition sur la dépendance faible du processus. Soient $\mathcal{F}_a^b$ la tribu engendrée par les variables aléatoires $(X_i)_{a\leq i\leq b}, -\infty \leq a \leq b \leq \infty$, et soit $L^2(\mathcal{F}_a^b)$ l'espace de toutes les variables aléatoires qui sont $\mathcal{F}_a^b$-mesurables, de variance finie. Le coefficient de corrélation maximale $\rho^*(n)$ entre $\{X_i, i \leq 0\}$ du passé et $\{X_i, i \geq n\}, n > 0$ du futur de cette suite stationnaire $(X_i, i \in \mathbb{Z})$ est défini par :

$$\rho^*(n) = \sup_{f\in L^2(\mathcal{F}_{-\infty}^0),\ g\in L^2(\mathcal{F}_n^\infty)} \frac{Cov(f,g)}{\sqrt{Var(f)Var(g)}}.$$

Nos résultats principaux sont les suivants :

- `Dans le cas à temps discret`

Notre premier resultat est la Loi (forte) de Grandes Nombres pour les estimateurs $D_n(j)$ et $\bar{D}_n(j)$.

Théorème 0.36 *Soit* $\{X_i, i \geq 0\}$ *un processus gaussien stationnaire et ergodique, de distribution marginale normale* $\mathcal{N}(0, \sigma^2)$. *Alors,*

$$\lim_{n\to\infty} D_n(j) = \lim_{n\to\infty} \bar{D}_n(j) = \mathbb{E}(X_j|\Delta),\ p.s.$$

En statistiques, le TCL joue un rôle central. Il est l'objetif du résultat suivant.

Théorème 0.37 *Soit* $\{X_i, -\infty \leq i \leq \infty\}$ *un processus gaussien stationnaire et ergodique, de distribution marginale normale* $\mathcal{N}(0, \sigma^2)$. *Si*

$$\sum_{n=0}^{\infty} \rho^*(n) < +\infty, \tag{0.42}$$

alors,

$$\sqrt{n}\,(D_n(j) - \mathbb{E}(X_j|\Delta)) \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}^2)$$
$$\sqrt{l(n)}\left(\bar{D}_n(j) - \mathbb{E}(X_j|\Delta)\right) \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}^2)$$

où

$$\begin{aligned}\tilde{\sigma}^2 = \frac{1}{P(\Delta_0)}&\Big[Var\left(1_{\Delta_0}\left(X_j - \mathbb{E}(X_j|\Delta_0)\right)\right) \\ &+2\sum_{k=1}^{+\infty} Cov\left(1_{\Delta_0}\left(X_j - \mathbb{E}(X_j|\Delta_0)\right), 1_{\Delta_k}\left(X_{k+j} - \mathbb{E}(X_{k+j}|\Delta_k)\right)\right)\Big].\end{aligned} \tag{0.43}$$

où $\Delta_k = \{\omega; (X_k, \cdots, X_{k+d-1}) \in \Delta\}$ *; la dernière série étant absolument convergente.*

- `Dans le cas à temps continu`

Soit $(X_t)_{t\geq 0}$ un processus centré gaussien stationnaire et ergodique dont le coefficient de corrélation $\rho(t) := \rho(X_0, X_t)$ est continu. Nous le prolongeons sur $\mathbb{R}$ par $\rho(t) := \rho(-t)$ pour tout $t < 0$. Par le Théorème de Bochner, il existe une mesure positive et **symétrique** bornée μ sur $\mathbb{R}$ telle que

$$\sigma^2\rho(t) = \int_{\mathbb{R}} e^{itx} d\mu(x), \ \forall t \in \mathbb{R}.$$

La mesure μ s'appelle **mesure spectrale** de (X_t).

Puisque $\mathbb{E}(X_t|X_0 = a) = a\rho(t)$ et que $\{X_0 \leq a, X_h \geq a\}$ devrait approcher l'événement $X_0 = a$ intuitivement, du fait de la continuité de $\rho(t)$ lorsque $h \to 0_+$(on peut même supposer que $\rho(t)$ est Hölderien, ce qui impliquerait la continuité des trajectoires du processus (X_t) (ou plutôt d'une version de celui-ci) par le théorème de Kolmogorov), donc il est tout à fait légitime d'imaginer que

$$\lim_{h\to 0+} \mathbb{E}(X_t|X_0 \leq a, X_h \geq a) = a\rho(t).$$

Le fameux paradoxe de Kac-Slepian [55] dit très exactement que ce genre d'intuition pourrait être fausse.

L'EDA pour estimer $a\rho(t)$ $(t > 0)$ est défini par

$$D_n(h,t) = \frac{1}{n}\sum_{k=1}^{n} X_{\tau_k(h)+t}$$

où $\tau_k(h)$ sont les moments successifs de (X_{kh}) en montant le niveau a, et la maille $h > 0$ est petite telle que $t/h \in \mathbb{N}^*$. On définit $\bar{D}_n(h,t)$ de la même manière que $\bar{D}_n(j)$. Par le Théorème 0.36, on a presque sûrement,

$$\lim_{n\to\infty} D_n(h,t) = \lim_{n\to\infty} \bar{D}_n(h,t) = \mathbb{E}\left(X_t|X_0 \leq a, X_h \geq a\right).$$

Théorème 0.38 *Soit $(X_t)_{t\geq 0}$ un processus centré gaussien ergodique et stationnaire tel que $\rho(t) := \rho(X_0, X_t)$ est continue sur $\mathbb{R}^+$ et dérivable pour $t > 0$.*

(a) Si

$$\Gamma = \int_{\mathbb{R}} x^2 d\mu(x) < +\infty \ (\Leftrightarrow \rho''(0) > -\infty)$$

alors, on a presque sûrement,

$$\lim_{h=T/m\to 0+}\lim_{n\to\infty} D_n(h,t) = \rho(t)a - \sqrt{\frac{\pi}{2\Gamma}}\rho'(t)\sigma^2 = \rho(t)a - \sqrt{\frac{\pi}{2}}\frac{\rho'(t)\sigma}{\sqrt{-\rho''(0)}}$$

(b) Si $\Gamma = \int_{\mathbb{R}} x^2\mu(dx) = +\infty$*, alors on a presque sûrement*

$$\lim_{h=T/m\to 0+} \lim_{n\to\infty} D_n(h,t) = \rho(t)a.$$

En plus, les assertions (a) et (b) sont encore vraies pour $\bar{D}_n(h,t)$.

Remarquons que $\sigma^{-2}\mu(dx)$ est la distribution de la fréquence des bruits. Donc ce théorème montre que $D_n(h,t)$, et $\bar{D}_n(h,t)$ sont biaisés si le processus ne contient pas suffisamment de bruits de haute fréquence.

Un exemple important en mécanique est l'oscillateur harmonique excité par un bruit blanc, décrit par

$$\begin{cases} dX_t = V_t dt \\ dV_t = \varsigma dW_t - (2\kappa\omega_0 V_t + {\omega_0}^2 X_t)dt \end{cases} \tag{0.44}$$

où $\omega_0 > 0$ est la pulsation constante, et $\kappa > 0$ est le coefficient d'amortissement. A l'aide du Théorème 0.38 et d'un calcul assez laborieux, on montre que pour (X_t), l'EDA $D_n(h,t)$ ou $\bar{D}_n(h,t)$ est consistent si $\kappa \neq 1$ (la zone de non-résonance). Dans le cas de résonance (où $\kappa = 1$), l'EDA possède un biais asymptotique :

$$\lim_{h=t/m\to 0} \lim_{n\to\infty} D_n(h,t) = \rho(t)a + \sqrt{\frac{\pi\varsigma^2}{8\kappa{\omega_0}^3}} te^{-\omega_0\kappa t}.$$

Bibliography

[1] Adams, T.M. and Nobel, A.B. (1998) On density estimation from ergodic processes. *Ann. Probab.*, **26**, 794-804.

[2] Aleshkyavichene, A.K. (1990) Probabilities of large deviations for U-statistics and von Mises functionals. *Teor. Veroyatnost. i Promenen*, **35**, 3-14.

[3] Bahadur, R.R. and Zabell, S.L. (1979) Large deviations of the sample mean in general vector space, *Ann. Probab.*, **7**, 587-621.

[4] Ben Arous, G. and Brunaud, M. (1990) Methode de Laplace: etude variationnelle des fluctuations de diffusions de type "Champs moyens". *Stochastics,* **31**,79-144.

[5] Ben Arous, G. and Guionnet, A. (1995) Large deviations for Langevin spin glass dynamics. *Prob. Th. Rel. Fields*, **102**, 455-509.

[6] Bernard, P. (1988) Identification de grandes structures: une remarque sur la méthode du décrément aléatoire. *Journal de Mécanique théorique et appliquée*, **7**, 3, 269-280.

[7] Billingsley, P. (1968) *Convergence of Probability Measures*, John Wiley and Sons Inc.

[8] Bobkov, S.G. and Götze, F. (1999) Exponential Integrability and Transportation Cost Related to Logarithmic Sobolev Inequalities. *J. Funct. Anal.*, **163**, 1-28.

[9] Bosq, D., Merlevède, F., Peligrad, M. (1999) Asymptotic normality for density kernel estimators in discrete and continuous time. *J. Multivariate Anal.*, **68**, 78-95.

[10] Bramson, M., Cox, J.T. and Griffeath, D. (1988) Occupation time large deviations of the voter model. *Prob. Th. Rel. Fields*, **77**, 401-413.

[11] Bryc, B. and Dembo, A. (1996) Large deviations and strong mixing. *Ann. Inst. H. Poincaré. (Série Proba. et Stat.)*, **32**, 549-569.

[12] Bucklew, J.A. (1990) *Large deviation techniques in decision, simulation, and estimation*, New York: Wiley.

[13] Bucklew, J.A. (2004) Large deviation results for U-statistics. *Introduction to Rare Event Simulation*, series: Spring Series in Statistics.

[14] Carrasco, M. and Chen, X. (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, **18**, 17-39.

[15] Caughey, T.K. and Stumpf, H.J. (1961) Transient response of dynamical system under random excitation. *Jal of Applied Mechanics*, **4**, 563-566.

[16] Chaganty, N.R. and Sethuraman, J. (1987) Limit theorem in the area of large deviations for some dependent random variables. *Ann. Probab.*, **15**, 628-645.

[17] Chernoff, H. (1952) A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.*, **23**, 493-507.

[18] Chung, K.L. (1974)*A course in probability theory.* Second Edition, Academic Press New York and London.

[19] Cole, H.A. (1973) On-line failure Detection and damping Measurement of Aerospace Structures by Random Decrement Signatures. NASA CR, 2205.

[20] Cox, J.T. and Durret, R. (1990) Large deviations for independent random walks. *Prob. Th. Rel. Fields*, bf 84, 67-82.

[21] Cramér, H. (1938) Sur un nouveau théorème-limite de la théorie des probabilités. In *Actualités Scientifiques et Industrielles*, number 736 in Colloque consacré à la théorie des probabilités, pages 5-23. Hermann, Paris.

[22] Csörgö, M. and Horváth, L. (1988) Central limit theorems for L_p-norms of density estimators. *Z. Wahrsch. Verw. Gebiete.*, **80**, 269-291.

[23] Dembo, A. and Shao, Q.M (1998) Self-normalized large deviations in vector space. In: *Progress in Probability*, (Eberlein, Hahn, Talagrand, Eds), **43**, 27-32.

[24] Dembo, A. and Shao, Q.M. (1998) Self-normalized moderate deviations and LILs. *Stochastic Process. Appl.*, *75*, 51-65.

[25] Dembo, A. and Zeitouni, O. (1998) *Large Deviations Techniques and Applications.* 2nd Edn, Springer, New York.

[26] Deuschel, J.D. (1989) Invariance principle and empirical mean large deviations of the critical Orenstein-Uhlenbeck process. *Ann. Probab.*, **17**, 74-90.

[27] Deuschel, J.D. and Stroock, D.W. (1989) *Large Deviations.* Pure Appl. Math., vol 137, Academic Press, San Diego.

[28] Devroye, L. (1983) The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates. *Ann. Statist.*, **11**, 896-904.

[29] Devroye, L. (1988) The Kernel Estimate is Relatively Stable *Prob. Theo. Rela. Field.*, **77**, 537-550.

[30] Djellout, H., Guillin, A. and Wu, L.M. (2004) Transportation cost-information inequalities for random dynamical systems and diffusions. *Ann. Probab.*, **32**, 2702-2732.

[31] Dobrushin, R.L., Kotechy, R. and Shlosman, S. (1992) Wulff construction: a global shape from local interactions. AMS Translation Series, providence.

[32] Donsker, M.D. and Varadhan, S.R.S. (1975; 1983) Asymptotic evaluation of certain Markov process expectations for large time, I-IV. *Commun. Pure Appl. Math.*, (1975) **28**, 1-47, 279-301; **29** 389-461; (1983) **36** , 183-212.

[33] Dupuis, P. and Ellis, R.S. (1995) The large deviation principle for a general class of queueing systems. *I. Trans. Amer. Math. Soc.* , **347**, 2689-2751.

[34] Durrett, R. (1985) Particle Systems, Random Media and Large Deviations. *Contemp. Math. 41.* Amer. Math. Soc., Providence, RI.

[35] Durrett, R. and Schonmann, R.H. (1988) Large deviations for the contact process and two demensional percolation. *Prob. Th. Rel. Fields* , **77**, 583-603.

[36] Ellis, R.S. (1985) *Entropy, Large deviations and Statistical Mechanics*, Springer-Verlag, New York.

[37] M.I. Freidlin and A.D. Wentzell, *Random Perturbation of Dynamical Systems*, translated by J. Szuc, Springer, Berlin 1984.

[38] Gamboa, F and Gassiat, E. (1997) Bayesian methods and maximum entropy for ill-posed inverse problems. *Ann. Stat.*, **25**, 328-350.

[39] Gao, F.Q. (2003) Moderate deviations and large deviations for kernel density estimators. *J. Theoret. Probab.*, **16**, 401-418.

[40] Gao, F.Q., Moderate deviations and the law of iterated logarithm for kernel density estimators in L_1. Preprint 2003.

[41] Gong, F.Z. and Wu, L.M., (2000) spectral gap of positive operators and applications. *C. R. Acad. Sci. Paris Sér. I Math.*, **331**, 983-988.

[42] Giné, E., Mason, D.M. and Zaitsev, A.Yu. (2003) The L_1-norm density estimator process. *Ann. Probab.*, **31**, 719-768.

[43] Giné, E. and Mason, D.M. (2004) The law of the iterated logarithm for the integrated squared deviation of a kernel density estimator. *Bernoulli*, **10**, 721-752.

[44] Giné, E., Koltchinskii, V. and Zinn, J. (2004) Weighted uniform consitency of kernel density estimators. *Ann. Probab.*, **32**, 2570-2605.

[45] Hall, P. and Heyde, C.C. (1980) *Martingale limit theory and its application*, Academic Press INC, New York.

[46] Hoeffding, W. (1965) On probabilities of large deviations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probabilty*, 203-219, Berkeley. University of California Press.

[47] Holmes, R.B. (1975) *Geometric Functional Analysis and its Applications*, G.T.M. 24, Springer-Verlag.

[48] Horváth, L. (1991) On L_p-norms of multivariate density estimators. *Ann. Statist.*, **19**, 1933-1949.

[49] Horváth, L. and Shao, Q.M. (1996) Large deviations and laws of the integrated logarithm for partial sums normalized by the largest absolute observation. *Ann. Probab.*, *24*, 1368-1387.

[50] Ibragimov, I.A. and Linnik, Y.V. (1971) *Independent and stationary sequences of variables random* , Ed. by J.F.C. Kingman, Wolters-Noordhoff, Groningen.

[51] Ibrahim, S.R. and Mikulcik, E.C. (1977) A Method for ther Direct Identification of Vibration Parameters from the Free Response. *The Shock and Vibration Bulletin*, **47**, 183-198.

[52] Ioffe, D. (1995) Exact large deviation bounds up to T_c for the Ising model in two demensions. *Prob. Th. Rel. Fields*, **102**, 313-330.

[53] Jing, B.Y., Shao, Q.M. and Wang, Q.Y. (2003) Self-normalized Cramér type large deviations for independent random variables. *Ann. Probab.*, **31**, 2167-2215.

[54] Joutard, C. (2004) Grandes déviations en statistique asymptotique. Thèse, Université Toulouse III - Paul Sabatier.

[55] Kac,M. and Slepian, D. (1959) Large excursions of Gaussian processes. *Annals of Math Stat*, **30**, 1215-1228.

[56] Kato, T. (1992) *Perturbation Theory for Linear Operators.* 2nd Edn, Spring-Berlin, New York/Berlin.

[57] Kipnis, C. and Olla, S. (1990) Large deviations for m the hydrodynamical limit for a system of independent Brownian particles. *Stochatstics and Stoch. Rep.*, **33**, 17-25.

[58] Kolmogorov A. N. and Rozanov Yu. A. (1960) On strong mixing conditions for stationary Gaussian processes. *Theory of Probability and Its Applications*, **5**, 204-208.

[59] Kristensen, D. (2004) Lecture notes on Economics Methods, chapter 7, Department of Economics, University of MADISON, WISCONSIN.

[60] Kusuoka, S. and Tamura, Y. (1984) The convergence of Gibbs measures Associated with mean fields potentials. *J. Fac. Sci. Univ. Tokyo*, **31**, 223-245.

[61] Bebowitz, J.L. and Schonmann, R.H. (1988) Pseudo free energies and large deviations for non Gibssian FKG measures. *Prob. Th. Rel. Fields*, **77**, 49-64.

[62] Ledoux, M. (1999) Concentration of measure and logarithmic Sobolev inequalities, *Séminaire de Probab. XXXIII*, LNM **1709**, 120-216, Springer.

[63] Lei, L.Z., Wu, L.M. and Xie, B. (2003) Large deviations and deviation inequality for kernel density estimator in $L_1(R^d)$-distance. In *Development of Modern Statistics and Related Topics*, Ser. Biostat. 1., 89-97, *World Sci. Publishing, River Edge, NJ.*

[64] Lei, L.Z. and Wu, L.M. (2004) The exponential convergence of kernel density estimator in L_1 for ϕ-mixing processes. *Ann.I.S.U.P.*, **48**, 59-68.

[65] Lei, L.Z. and Wu, L.M. (2005) Large deviations of kernel density estimator in L_1 for uniformly ergodic Markov processes. *Stochastic Process. Appl.*, **115**, 275-298.

[66] Leonard, C. (1987) Large deviations and law of large numbers for a mean field type interacting particle system. *Stochastic Process. Appl.*, **25**, 215-235.

[67] Louani, D. (1998) Large deviations limit theorems for the kernel density estimator. *Scan. J. Stat*, **25**, 243-253.

[68] Louani, D. (2000) Large deviations for the L_1-distance in kernel density estimation. *J. Statist. Plann. Inference.*, 177-182.

[69] Marton, K. (1996) A measure concentration inequality for contracting Markov chains.*Geom.Funct.Anal.*, **7**. 556-571.

[70] Menneteau, L. Large deviations for local empirical measures. High dimensional probability, II (Seattle, WA, 1999), 293–312, Progr. Probab., 47, Birkhuser Boston, Boston, MA, 2000.

[71] Meyn, S.P. and Tweedie, R.L. (1993) *Markov chains and stochastic stability.* Springer, Berlin.

[72] Nakayama, Y., Watanabe, T. and Fujisaka, H. (2001) Self-similar fluctuation and large deviation statistics in the shell model of turbulence. *Phys ReV E Stat Non Lin Soft Matter Phys*, Nov, 64(5 pt 2).

[73] Kontoyiannis, I and Meyn, S.P. (2003) Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.*, **13**, 304-362.

[74] Orey, S. (1988) Large deviations for the empirical field of Curie-Weiss models. *Stochastics*, **25**, 3-14.

[75] Papangelou, F. (1990) Large deviations and the internal fluctuations of critical mean field systems. *Stochastic Process. Appl.*, **36**, 1-14.

[76] Parzen, E. (1962) On the estimation of a probabilty density function and the mode. *Ann. Math. Statist.*, **33**, 1065-1076.

[77] Peligrad, M. (1992) Properties of uniform consistency of the kernel estimators of density and of regression functions under dependence assumption. *Stochastics Stochastics Rep.*, **40**, 147-168.

[78] Pfister, C. (1991) Large deviations and phase separation in the two dimensional Ising model. *elv. Phys. Acta*, **64**, 953-1054.

[79] Pisztora, A. (1996) Surface order large deviations for Ising, Potts and percolation models. *Prob. Th. Rel. Fields*, **104**, 427-466.

[80] Ruelle D. (1978) *Thermodynamic Formalism*, Addison-Wesley.

[81] Rio, E. (2000) Inégalités de Hoeffding pour les fonctions lipchitziennes de suites dépendantes. *CRAS, Paris, t.330, Série I*, 905-908.

[82] Samson, P.M. (2000) Concentration of measure inequalites for Markov chains and ϕ-mixing processes. *Ann. Probab.*, **28**, 416-461.

[83] Sarmanov, O.V. (1958) Maximal coefficient of correlation (nonsymmetric case), *Doklady Akademii Nauk SSSR*, **121**, no. 1 (in Russian).

[84] Shao, Q.M. (1997) Self-normalized large deviations, *Ann. Probab.*, **25**, 285-328.

[85] Shao, Q.M. (1999) A Cramér type large deviation result for Student's t-statistic. *J. Theoret. Probab.*, **12**, 387-398.

[86] Schonmann, R. (1987) Second order large deviation estimators for ferromagnetic systems in the phase coexistence region. *Comm. Math. Physics*, **112**, 409-422.

[87] Schonmann, R.H. and Shlosman, S.B. (1995) Complete analyticity for 2D Ising model completed. *Comm. Math. Physics*, **170**, 453-482.

[88] Silverman, B. W. (1978) Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27**, 832-835.

[89] Stroock, D.W. and Zegarlinski, B. (1992) The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Comm. Math. Physics*, *149*, 175-193.

[90] Watanabe, T., Nakayama, Y. and Fujisaka, H. (2000) Large deviation statistics of the energy-flux fluctuation in the shell model of trubulence. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*, Feb 2000; **61**(2): R1024-7.

[91] Worms, J. (2000) Principes de déviations modérées pour des martingales et applications statistiques. Thèse de Doctorat, Université de Marne-La-Vallée.

[92] Wu, L.M. (1995) Moderate deviations of dependent random variables related to CLT. *Ann. Probab.*, **23**, 420-445.

[93] Wu, L.M. (1997) An introduction to large deviations. In: Yan, J.A., Peng, S.G., Fang, S.Z. and Wu, L.M. (Eds), *Several Topics in Stochastic Analysis*, 225-336, Academic Press of China, Beijing, (In Chinese).

[94] Wu, L.M. (2000) Uniformly integrable operator and large deviations for Markov processes. *J. Funct. Anal.*, **172**, 301-376.

[95] Wu, L.M. (2000b) A deviation inequality for non-reversible Markov processes. *Ann. Inst. H. Poincaré Probab. Statist.*, **36**, 435-445.

[96] Wu, L.M. (2001) Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems, *Stochastic Process. Appl.*, **91**, 205-238.

[97] Wu, L.M. (2002) Large deviations for reversible Markov chains. Preprint.

[98] Wu, L.M. (2005) A Φ-entropy contraction inequality for Gaussian vectors and applications to Gaussian Markov processes. Preprint 2005.

[99] Zeitouni, O. and Gutman, M. (1991) On universal lypotheses testing via large deviations. *IEEE Trans. Inf. Theory*, IT-**37**, 285-290.

[100] Zeitouni, O. and Zakai, M. (1992) On the optimal tracking problem. SIAM J. *cont. Opt.*, **30**, 426-439. Erratum, **32**, 1194, (1994).

(Une partie des réferences ci-dessus ne sont pas citées dans cette Introduction, mais utilisées dans les chapitres qui suivent.)

Chapter 1

Large deviations and deviation inequality for kernel density estimator in $L^1(\mathbb{R}^d)$-distance

(Published in: *Development of Modern Statistics and Related Topics***)**

This article is a joint work with Professor Liming Wu and Bin Xie.

1.1 Introduction

Let $\{X_i; i \geq 1\}$ be a sequence of independent and identically distributed random variables (i.i.d.r.v in short), taking values in $\mathbb{R}^d$, defined on probability space $(\Omega, \mathcal{F}, P)$ with distribution measure $d\mu = f(x)dx$, where the density $f \in L^1(\mathbb{R}^d)$ is unknown. The empirical measure is $L_n = \dfrac{1}{n}\sum_{i=1}^n \delta_{X_i}$. Let K be a measurable function such that

(H1) $K \geq 0,\ \int_{\mathbb{R}^d} K dx = 1.$

and set $K_h(x) = \dfrac{1}{h^d}K(\dfrac{x}{h})$. The kernel density estimator of f is defined as:

$$f_n^*(x) = K_{h_n} * L_n = \frac{1}{n}\sum_{i=1}^n \frac{1}{h_n^d} K\left(\frac{x - X_i}{h_n}\right), \quad x \in \mathbb{R}^d \tag{1.1}$$

where $\{h_n, n \geq 1\}$ is the bandwidth, that is, a sequence of positive numbers satisfying

(H2) $h_n \to 0, \quad nh_n^d \to +\infty \quad as \quad n \to \infty.$

The limit behavior of f_n^* in $(L^1(\mathbb{R}^d), \|\cdot\|_1 := \|\cdot\|_{L^1(\mathbb{R}^d)})$ is a subject of current study. L. Devroye, in a fundamental paper [2] on the subject, proved under (H1) and (H2), that $\|f_n^* - f\|_{L^1(\mathbb{R}^d)} \to 0,\ a.s.$ (the strong consistence), and proved the following exponential convergence: for each $r > 0$, there are $C, \delta > 0$ such that

$$\mathbb{P}\left(\|f_n^* - f\|_1 > r\right) \le Ce^{-\delta n},\ \forall n \ge 1 \tag{1.2}$$

More recently Louani [7] obtains the large deviation estimation associated with (1.2), i.e., giving an exact identification of δ as $n \to \infty$. To state his result, define, for any $0 \le a \le 1$, the following functions:

$$\Gamma_a^+(r) := \begin{cases} (a + \frac{r}{2}) \log(1 + \frac{r}{2a}) + (1 - a - \frac{r}{2}) \log(1 - \frac{r}{2(1-a)}) & \text{if } 0 < r < 2 - 2a \\ +\infty, & \text{otherwise} \end{cases} \tag{1.3}$$

$$\Gamma_a^-(r) := \begin{cases} (a - \frac{r}{2})\log(1 - \frac{r}{2a}) + (1 - a + \frac{r}{2}) \log(1 + \frac{r}{2(1-a)}) & \text{if } 0 < r < 2a \\ +\infty, & \text{otherwise} \end{cases} \tag{1.4}$$

$$\begin{aligned} \Gamma_a(r) &:= \min\{\Gamma_a^+(r), \Gamma_a^-(r)\}; \\ l(r) &:= \inf\{\Gamma_a(r) : 0 \le a \le 1\} \end{aligned} \tag{1.5}$$

Now the main result of Louani [7] is stated as

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\|f_n^* - f\|_1 > r\right) = -l(r),\ \forall r > 0. \tag{1.6}$$

To complement those two results, we study in this paper the following two questions:

Question 1. What is the large deviation estimation of $\mathbb{P}\left(\|f_n^* - g\|_1 \le \delta\right)$ for any given density function g ?

Question 2. For each fixed n and deviation $r > 0$, how can one bound $\mathbb{P}\left(\|f_n^* - f\|_1 > r\right)$?

For large deviations on $L^\infty(\mathbb{R}^d)$ which is much more difficult, the reader is referred to Gao [4] and the references therein.

1.2 Main results

For the language of large deviation, the reader is referred to [1], [8]. Our first result may seem a little abstract. It extends the well known Sanov theorem ([1], [8]).

Proposition 1.1 *Assume* **(H1)** *and $h_n \to 0$ (without* **(H2)***). Then as n goes to infinity, $\mathbb{P}(f_n^* \in \cdot)$ satisfies the large deviation principle (LDP in short) on $L^1(dx)$ w.r.t. the weak topology $\sigma(L^1, L^\infty)$), with the rate function given by*

$$I(g) = \begin{cases} Ent_\mu(g) := \int_{\mathbb{R}^d} g(x) \log \frac{g(x)}{f(x)} dx, & \text{if } g \in \mathcal{P}, g(x)dx \ll f(x)dx; \\ +\infty & \text{otherwise} \end{cases} \tag{1.7}$$

where $\mathcal{P}$ is the set of all probability density functions on $\mathbb{R}^d$. More precisely

(GRF) I is a Good Rate Function on $(L^1(\mathbb{R}^d), \sigma(L^1, L^\infty))$, i.e., for any $L \geq 0$, $[I \leq L]$ is compact in $(L^1(\mathbb{R}^d), \sigma(L^1, L^\infty))$.

(LLD) (Lower bound of Large Deviation) For any open subset $G \subset (L^1(\mathbb{R}^d), \sigma(L^1,$-$L^\infty))$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(f_n^* \in G\right) \geq - \inf_{g \in G} I(g).$$

(ULD) (Upper bound of Large Deviation) For any closed subset $F \subset (L^1(\mathbb{R}^d), \sigma(L^1,$-$L^\infty))$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(f_n^* \in F\right) \leq - \inf_{g \in F} I(g).$$

Remark: Because I is not a good rate function on$L^1(\mathbb{R}^d)$ w.r.t. the norm $\|\cdot\|_1$-topology, $\mathbb{P}(f_n^* \in \cdot)$ does not satisfy the LDP on $(L^1(\mathbb{R}^d), \|\cdot\|_1)$.

The following result says however that $\mathbb{P}(f_n^* \in \cdot)$ satisfies the weak*-LDP on $(L^1(\mathbb{R}^d), \|\cdot\|_1)$. It gives a satisfactory answer to Question 1) in the Introduction.

Theorem 1.2 *Assume* **(H1)** *and* **(H2)***. Then for any $g \in L^1(\mathbb{R}^d)$,*

$$\begin{aligned} &\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\|f_n^* - g\|_{L^1(\mathbb{R}^d)} < \delta) \\ &= \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\|f_n^* - g\|_{L^1(\mathbb{R}^d)} < \delta) = -I(g). \end{aligned} \tag{1.8}$$

Moreover for any $\|\cdot\|_1$-open and **convex** *subset G in $L^1(\mathbb{R}^d)$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(f_n^* \in G) = - \inf_{g\in G} I(g). \tag{1.9}$$

Remark: By (1.8), for any $\|\cdot\|_1$-open subset G in $L^1(\mathbb{R}^d)$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(f_n^* \in G) \geq - \inf_{g\in G} I(g).$$

In particular, we have for any $r > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\|f_n^* - f\| > r) \geq -\inf\{I(g);\ \|g-f\|_1 > r\}.$$

We can prove that $\inf\{I(g);\ \|g-f\|_1 > r\} = l(r)$, is exactly the rate function (1.5) found by Louani [7]. In other words the lower bound here is much more general.

We now present our answer to Question 2.

Theorem 1.3 *Assume* **(H1)** *and . Let* $J_n := \int_{\mathbb{R}^d} |f_n^*(x) - f(x)|dx = \|f_n^* - f\|_1$. *Then for any* $n \geq 1$ *and* $r > 0$,

$$\mathbb{P}(|J_n - \mathbb{E}J_n| > r) \leq 2\exp\left(-\frac{nr^2}{8}\right). \tag{1.10}$$

As $\mathbb{E}J_n \to 0$ under **(H1)** and **(H2)** ([2]), the inequality above is much more precise than (1.2). About the deviation inequality L.Devroye has in his paper [3] got under **(H1)**,

$$\mathbb{P}(|J_n - \mathbb{E}J_n| > \epsilon) \leq 2\exp -\frac{n\epsilon^2}{32 \times 32}$$

but only for ε small. His method is based on Poissonisation of n. Our proof will be completely different and based on the known deviation inequality of product measure w.r.t. the Hamming Distance, see the excellent monograph by Ledoux [6].

1.3 Proofs of the main results

1.3.1 Proof of Proposition 1.1

Step 1 (identification of the Cramér function). For any $k \in L^\infty = (L^1)^*$, we calculate the Cramér functional by means of the i.i.d. property,

$$\begin{aligned}
\Lambda(k) &:= \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\exp\left(n\langle f_n^*, k\rangle_{dx}\right) \\
&= \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\exp\left(\sum_{i=1}^n \frac{1}{h_n^d} \int K\left(\frac{x-X_i}{h_n}\right) k(x)dx\right) \\
&= \lim_{n\to\infty} \log \mathbb{E}\exp\left(\frac{1}{h_n^d} \int K\left(\frac{x-X}{h_n}\right) k(x)dx\right)
\end{aligned}$$

where $X = X_1$. But it is well known that $\frac{1}{h_n^d} \int K\left(\frac{x-y}{h_n}\right) k(x)dx$ tends to $k(y)$ in measure dy. Since the law of X is absolutely continuous, then $\frac{1}{h_n^d} \int K\left(\frac{x-X}{h_n}\right) k(x)dx$

converges to $k(X)$ in $\mathbb{P}$-probability. Moreover, $|\frac{1}{h_n^d}\int K\left(\frac{x-X}{h_n}\right)k(x)dx| \le \|k\|_\infty$. Consequently by dominated convergence, we obtain

$$\begin{aligned}\Lambda(k) &= \log \mathbb{E}\exp(k(X)) \\ &= \log \int e^{k(x)}\mu(dx)\end{aligned} \tag{1.11}$$

Moreover by the famous variational formula of entropy of Donsker-Varadhan, the Legendre transformation $\Lambda^*:\ L^1(\mathbb{R}^d) \to [0,+\infty]$ of Λ is given by

$$\begin{aligned}\Lambda^*(g) :&= \sup\{\langle g,k\rangle_{dx} - \Lambda(k);\ k \in L^\infty\} \\ &= \sup\{\langle g,k\rangle_{dx} - \log\int e^{k(x)}\mu(dx);\ k \in L^\infty\} \\ &= \begin{cases} Ent_\mu(g), & \text{if } \ g \in \mathcal{P}; \\ +\infty & \text{otherwise}\end{cases}\end{aligned}$$

In other words $\Lambda^*(g) = I(g)$ given in (1.7).

Step 2: LLD. Since $\Lambda(k)$ is Gateaux-differentiable on $L^\infty(\mathbb{R}^d)$, hence the LLD in (ii) follows by the abstract Ellis-Gärtner theorem (see [8], Theorem 2.7).

Step 3: GRF + ULD. Again by the abstract Ellis-Gärtner theorem (see [8], Theorem 2.1), it is enough to show that if $g: k \to \langle g,k\rangle$ is a linear form on $L^\infty(\mathbb{R}^d)$ such that

$$\bar{\Lambda}^*(g) := \sup\{\langle g,k\rangle - \Lambda(k);\ k \in L^\infty\} < +\infty,$$

then $g \in L^1(\mathbb{R}^d)$. By following the proof of [8], Theorem 3.3, it suffices to prove that $\Lambda(k_n) \to 0$ for any sequence $(k_n)_{n\ge 0}$ of functions in $L^\infty(\mathbb{R}^d)$ decreasing $dx - a.e.$ to zero. The last property follows by the explicit expression (1.11) of $\Lambda(k_n)$ and dominated convergence, for $e^{\|k_0\|_\infty} \ge e^{k_n(x)} \downarrow 0$, $\mu - a.e.$. □

1.3.2 Proof of Theorem 1.2.

The following basic result of Devroye is crucial.

Lemma 1.4 *(due to Devroye [2]) Under* **(H1)** *and* **(H2)**,

$$\int_{\mathbb{R}^d} |f_n^* - f|dx \to 0,\ \mathbb{P} - a.s$$

We can now prove the lower bound in Theorem 1.2,

Lemma 1.5 *Under* **(H1)** *and* **(H2)**, *we have for any* $g \in L^1(\mathbb{R}^d)$,

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\|f_n^* - g\|_{L^1(\mathbb{R}^d)} < \delta) \ge -I(g), \qquad \forall \delta > 0 \tag{1.12}$$

Proof. We may assume that $I(g) < +\infty$ (trivial otherwise). Without loss of generality we assume that $(\Omega, \mathcal{F}, \mathbb{P}) = ((\mathbb{R}^d)^{\mathbb{N}^*}, \mathcal{B}^{\mathbb{N}^*}, \mu^{\otimes \mathbb{N}^*})$ and $X_i(\omega) = \omega_i, i \geq 1$ are the coordinates on the product space Ω. Let $\mathcal{F}_n := \sigma(X_1, \cdots, X_n)$. The following method of transformation of measure is standard for treating the LLD. Since $I(g) < +\infty$, g is a probability density such that $\nu(dx) := g(x)dx \ll \mu(dx) = f(x)dx$ and $Ent_\mu(g) = \int \log(g(x)/f(x))d\nu(x) < +\infty$. Consider the product measure $\mathbb{Q} := \nu^{\otimes \mathbb{N}^*}$. Then

$$\mathbb{Q} \mid_{\mathcal{F}_n = \sigma(X_1, \ldots X_n)} = \prod_{i=1}^n \frac{g(X_i)}{f(X_i)} \mathbb{P} \mid_{\mathcal{F}_n}$$

We have for any $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}(\|f_n^* - g\|_1 < \delta) &\geq \int_{\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_n} > 0} \frac{d\mathbb{P}}{d\mathbb{Q}} \mid_{\mathcal{F}_n} \cdot 1_{[\|f_n^* - g\|_1 < \delta]} d\mathbb{Q} \\
&= \int_\Omega \exp\left(- \sum_{i=1}^n \log \frac{g}{f}(X_i) \right) \cdot 1_{[\|f_n^* - g\|_1 < \delta]} d\mathbb{Q} \\
&\geq \exp\left(-n[Ent_\mu(g) + \varepsilon]\right) \mathbb{Q}\left(A_{n,\varepsilon} \bigcap [\|f_n^* - g\|_1 < \delta] \right)
\end{aligned}$$

where $A_{n,\varepsilon} := \left[\frac{1}{n} \sum_{i=1}^n \log(g/f)(X_i) \leq Ent_\mu(g) + \varepsilon \right]$.

Now to prove (1.12), it remains to show that $\mathbb{Q}(A_{n,\varepsilon}) \to 1$ and $\mathbb{Q}([\|f_n^* - g\|_1 < \delta])$-$\to 1$, as n goes to infinity (for any $\varepsilon > 0$). The first is easy, because by the law of large number,

$$\frac{1}{n} \sum_{i=1}^n \log \frac{g}{f}(X_i) \to \mathbb{E}^{\mathbb{Q}} \log \frac{g}{f}(X) = Ent_\mu(g), \ \mathbb{Q} - a.s.$$

The second, i.e., $\mathbb{Q}([\|f_n^* - g\|_1 < \delta]) \to 1$, follows by Lemma 1.4 (applied to (X_i) under $\mathbb{Q}$). $\square$

We now turn to give the

Proof of Theorem 1.2. By Lemma 1.5, we have already

$$\liminf_{\delta \to 0} \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\|f_n^* - g\|_{L^1(\mathbb{R}^d)} < \delta) \geq -I(g).$$

For the upper bound,note that $\bar{B}(g, \delta) := \{\tilde{g} \in L^1(\mathbb{R}^d); \ \|\tilde{g} - g\| \leq \delta\}$ is convex, closed w.r.t. $\|\cdot\|_1$, then it is closed w.r.t. $\sigma(L^1, L^\infty)$. Thus by the ULD in Proposition

1.1, we have

$$\begin{aligned}
&\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\|f_n^*-g\|_{L^1(\mathbb{R}^d)}<\delta)\\
&\quad\le \limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(f_n^*\in\bar{B}(g,\delta))\\
&\quad\le -\liminf_{\delta\to 0}\inf_{\tilde{g}\in\bar{B}(g,\delta)} I(\tilde{g}) = -I(g)
\end{aligned}$$

where the last equality follows from the lower semi-continuity of I (by the GRF in Proposition 1.1). Hence (1.8) is established. The proof of (1.9) is similar. For any $g\in G$, choosing some ball $B(g,\delta)=\{\tilde{g}\in L^1(\mathbb{R}^d);\ \|\tilde{g}-g\|<\delta\}\subset G$, we have by Lemma 1.5,

$$l(G):=\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(f_n^*\in G)\ge -I(g).$$

As $g\in G$ is arbitrary, we obtain $l(G)\ge -\inf_G I$. For the upper bound, note that the closure of the convex subset G w.r.t. $\|\cdot\|_1$ or w.r.t. $\sigma(L^1,L^\infty)$ is the same (a well known consequence of the Hahn-Banach theorem), denoted by $\bar{G}$. Consequently by Proposition 1.1, we have

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(f_n^*\in G)\le -\inf_{g\in\bar{G}} I(g).$$

Hence for (1.9), it remains to show that $\inf_{g\in\bar{G}} I(g)=\inf_{g\in G} I(g)$. For this purpose let $g_1\in\partial G:=\bar{G}\backslash G$. By a known result in Banach space theory (see [5], (11.1), p59), for any $g_0\in G$ fixed, $g_t:=tg_1+(1-t)g_0\in G$ for any $t\in(0,1)$. As I is convex on $L^1(\mathbb{R}^d)$ (for it is the Legendre transformation of Λ), $t\to I(g_t)$ is convex on $[0,1]$. Thus

$$I(g_1)\ge\lim_{t\uparrow 1} I(g_t)\ge\inf_{g\in G} I(g)$$

where the desired equality follows for $g_1\in\partial G$ is arbitrary. □

1.3.3 Proof of Theorem 1.3.

All is based on the following

Lemma 1.6 *Let $(E_i,\mathcal{B}_i,\mu_i), i=1,\cdots,n$, be arbitrary probability space, and let $\mathbb{P}=\mu_1\otimes\cdots\otimes\mu_n$ be a product measure on the product space $E^n=E_1\times\cdots\times E_n$. A generic point in E^n is denoted by $x=(x_1,\cdots,x_n)$. Then, for every real measurable F on E such that $|F(x)-F(y)|\le 1$ whenever $x=(x_1,\cdots,x_n)$ and $y=(y_1,\cdots,y_n)$ only differ by one coordinate (this is equivalent to say that the Lipschitz coefficiant of F w.r.t. the Hamming distance is not greater than one). Then*

$$\mathbb{P}(F\ge\mathbb{E}^{\mathbb{P}}F+r)\le\exp\left(-\frac{r^2}{2n}\right).$$

See Ledoux [6] (Section 3.1, p162) for presentation of this important result. We now prove Theorem 1.3. We now prove Theorem 1.3. Without loss of generality we may assume that $X_1, \cdots, X_n$ are coordinates $x_1, \cdots, x_n$ on the product space $(\Omega, \mathcal{F}, \mathbb{P}) = ((\mathbb{R}^d)^n, \mathcal{B}^n, \mu^{\otimes n})$. Put

$$F(x_1, \cdots, x_n) := \frac{n}{2} J_n(x_1, \cdots, x_n) = \frac{n}{2} \int_{\mathbb{R}^d} \left| \frac{1}{n} \sum_{i=1}^{n} K_h(z - x_i)dz - f(z) \right| dz$$

where $h = h_n$. For any $i = 1, \cdots, n$ and for any $x, y \in (\mathbb{R}^d)^n$ such that $x_j = y_j$ for all $j \neq i$, we have

$$|F(x) - F(y)| \leq \frac{1}{2} \int_{\mathbb{R}^d} |K_h(z - x_i) - K_h(z - y_i)| dz \leq 1.$$

Consequently by Lemma 1.6 (applied to F and $-F$),

$$\begin{aligned} \mathbb{P}(|J_n - \mathbb{E}J_n| > r) &\leq \mathbb{P}\left(F - \mathbb{E}F > \frac{nr}{2}\right) + \mathbb{P}\left(F - \mathbb{E}F < -\frac{nr}{2}\right) \\ &\leq 2\exp\left(-\frac{1}{2n}\left[\frac{nr}{2}\right]^2\right) = 2\exp\left(-\frac{nr^2}{8}\right), \end{aligned}$$

the desired (1.10). $\square$

Acknowledgement: We are grateful to Prof. Gao Fuqing for useful conversations and communications of papers on the subject. This research was supported by the National Natural Science Foundation of China.

Bibliography

[1] Dembo, A., Zeitouni, O. (1998) *Large Deviations Techniques and Applications.* Second Edition. Springer.

[2] Devroye, L. (1983) The Equivaluce of Weak, Strong and Complete Convergence in L_1 For Kernel Density Estimates. *Ann. Stast.*, *11*, 896-904.

[3] Devroye, L. (1988) The Kernel Estimate is Relatively Stable. *Prob. Theo. Rela. Field.*, *77*, 537-550.

[4] Gao, F.Q. (2001) Moderate Deviations and Large Deviations For the Kernel Density Estimators. Preprint.

[5] Holmes, R.B. (1975). *Geometric Functional Analysis and its Applications*, G.T.M. 24, Springer-Verlag.

[6] Ledoux, M. (1999) *Concentration of measure and logarithmic Sobolev inequalities,* Séminaire de Probab. XXXIII, LNM **1709**, 120-216, Springer.

[7] Louani, D. (2000) Large deviations for the L_1-distance in kernel density estimation. *J. Statist. Plann. Inference.*, 177-182.

[8] Wu, L.M.(1997). *An introduction to large deviation, in Several Topics in Stochastics Analysis* (J.A.Yan, S.G.Peng, S.Z.Fang, L.M.Wu,Eds), 225-336, Academic Press of China,Beijing.[In Chinese]

Chapter 2

The exponential convergence of kernel density estimator in L^1 for ϕ-mixing processes

(Published in: *Annales de L'I.S.U.P.*)

This article is a joint work with Professor Liming Wu.

2.1 Introduction

Let $\{X_i; i \geq 1\}$ be a sample taken from a ϕ-mixing process with values in $\mathbb{R}^d$, defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with marginal distribution measure $d\mu = f(x)dx$, where the density $f \in L^1(\mathbb{R}^d)$ is unknown.

The empirical measure is $L_n = \dfrac{1}{n}\sum_{i=1}^n \delta_{X_i}$. Let K be a measurable function such that

(H1) $K \geq 0, \int_{\mathbb{R}^d} K dx = 1,$

and set $K_h(x) = \dfrac{1}{h^d} K(\dfrac{x}{h})$. The kernel density estimator of f is defined as usually as:

$$f_n^*(x) = K_{h_n} * L_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h_n^d} K\left(\frac{x - X_i}{h_n}\right), \quad x \in \mathbb{R}^d \tag{2.1}$$

where $\{h_n, n \geq 1\}$ is a sequence of positive numbers (bandwidth) satisfying

(H2) $h_n \to 0, \quad nh_n^d \to +\infty \quad as \quad n \to \infty.$

A natural distance of f_n^* from the unknown f is its L^1 distance below,

$$D_n^* = \int_{\mathbb{R}^d} |f_n^*(x) - f(x)|dx. \tag{2.2}$$

In the independent and identically distributed (i.i.d in short) case, L. Devroye, in a fundamental paper [5], proved

Theorem 2.1 *Let (X_i) be i.i.d. and K a nonnegative Borel measurable function on $\mathbb{R}^d$ with $\int K(x)dx = 1$ (i.e.,* **(H1)***), then the following conditions are equivalent:*

(i) $D_n^* \to 0$ in probability as $n \to \infty$ (weak consistency);

(ii) $D_n^* \to 0$ almost surely as $n \to \infty$ (strong consistency);

(iii) $D_n^* \to 0$ exponentially as $n \to \infty$, i.e., for any $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(D_n^* > \delta) < 0; \tag{2.3}$$

(iv) $\lim_n h_n = 0$ and $\lim_n nh_n^d = \infty$ (i.e., (H2)).

Recently Louani [11] (2000) prove the existence of limit in (2.3) and identifies that limit $-I(\delta)$ (i.e., a large deviation principle). And we (together with B. Xie) [10] establish a weak large deviation principle of f_n^* in L^1 (and it is known that the *good* large deviation principle fails). Gao [8] (2002) establishes the large and moderate deviation principle for f_n^* in L^∞ under mild condition on K, f and (h_n).

The study of kernel density estimators in the dependent cases were realized by many people from different points of view : see Peligrad [13] (1992), T.M. Adams and A.B. Nobel [1] (1998), Bosq, Merlevède and Peligrad [3] (1999) and the references therein. For instance Peligrad [13] established the uniform consistency of f_n^* (i.e., in L^∞ instead of L^1) under weaker condition on the ϕ-dependence coefficient (ϕ_k) than that used in this Note (nevertheless his conditions on $K, f, (h_n)$ are much stronger). In T.M. Adams and A.B. Nobel [1] (1998), a general procedure to construct ergodic processes for which the kernel density estimator *fails to be weakly consistent under* **(H1)** *and* **(H2)** is exhibited.

Note however that in the ϕ-mixing case, how to extend those results of large and moderate deviations in [11], [8], [10] in the i.i.d. case is an interesting open question. This question is quite delicate because even for stationary Doeblin recurrent Markov chains (for which ϕ_k decays exponentially to zero), the large deviation principle about partial sums fails in general (see Bryc and Dembo [4]).

In this Note we will carry out a first step towards the large deviations of f_n^*, i.e., to establish the exponential convergence of f_n^* to f in $L^1(\mathbb{R}^d, dx)$ for ϕ-mixing sequences verifying $\sum_k \phi_k < +\infty$, under **(H1)** and **(H2)**. Moreover we will yield an exponential inequality of Hoeffding type. Our main tool is the Hoeffding type inequality established recently by Rio [14](2000).

2.2 Main results

We briefly recall what is meant by the terminology of ϕ-mixing process. Given a sequence of random variables $(X_i)_{i\geq 1}$ with values in $\mathbb{R}^d$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$. For two sub-$\sigma$-algebras $\mathcal{A}, \mathcal{B}$ in $\mathcal{F}$, define

$$\phi(\mathcal{A}, \mathcal{B}) = \sup\left\{\left|\mathbb{P}(V) - \frac{\mathbb{P}(U\bigcap V)}{\mathbb{P}(U)}\right|;\ U \in \mathcal{A}, \mathbb{P}(U) \neq 0, V \in \mathcal{B}\right\}.$$

Define for every integer k,

$$\phi_k = \sup_{m\geq 1}\left\{\phi(\sigma(X_1, \cdots, X_m), \sigma(X_{m+l}; l \geq k))\right\}.$$

Theorem 2.2 *Let $(X_i)_{i\in\mathbb{N}^*}$ be a stationary sequence of $\mathbb{R}^d$-valued r.v. with marginal law $\mu(dx) = f(x)dx$. Assume that*

$$S_\phi := \sum_{k=1}^{\infty} \phi_k < +\infty. \tag{2.4}$$

Let K be a nonnegative measurable function on $\mathbb{R}^d$ with $\int K(x)dx = 1$ (i.e. **(H1)***) and (h_n) a sequence of positive numbers verifying* **(H2)***. Then $D_n^* \to 0$ exponentially as $n \to \infty$, i.e.,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(D_n^* > \delta) < 0, \qquad \forall \delta > 0.$$

Theorem 2.3 *In the context of Theorem 2.2, assume (2.4) and* **(H1)** *for K. Then for every $n \geq 1$ and all $r > 0$,*

$$\mathbb{P}(|D_n^* - \mathbb{E}D_n^*| > r/\sqrt{n}) \leq 2\exp\left(-\frac{r^2}{2(1+2S_\phi)^2}\right). \tag{2.5}$$

2.3 Some deviation inequalities for ϕ-mixing sequences

All of this Note is based on the following Hoeffding type inequality established recently by E. Rio [14] (2000).

Lemma 2.4 *([14]) Let $f : E^n \to \mathbb{R}$ satisfy*

$$|f(x) - f(y)| \leq L \tag{2.6}$$

for all $x, y \in E^n$ verifying $\#\{i;\ x_i \neq y_i\} = 1$. Then $\forall \lambda > 0$,

$$\begin{aligned} &\mathbb{E} \exp \left[\lambda \left(f(X_1, \cdots, X_n) - \mathbb{E} f(X_1, \cdots, X_n)\right)\right] \\ &\quad \leq \exp \left(\frac{\lambda^2}{8} \cdot nL^2(1 + 2S_\phi)^2\right); \end{aligned} \tag{2.7}$$

and in particular $\forall t > 0$,

$$\mathbb{P}\left(f(X_1, \cdots, X_n) - \mathbb{E} f(X_1, \cdots, X_n) > t\right) \leq \exp\left(-\frac{2t^2}{nL^2(1+2S_\phi)^2}\right). \tag{2.8}$$

Indeed (2.8) is [14], Corollaire 1 and (2.7) is an immediate consequence of [14], Théorème 1 together with the proof of Corollaire 1.

Consider the Hamming distance on E^n:

$$d_H(x, y) := \#\{i;\ x_i \neq y_i\}.$$

Condition (2.6) is equivalent to

$$|f(x) - f(y)| \leq L d_H(x, y),\ \forall x, y \in E^n$$

i.e., the Lipchitzian coefficient of f w.r.t. the Hamming distance d_H is less than L. In such way we can translate Rio's inequality (2.7) into the following transportation inequality, by Bobkov-Götze [2]:

Corollary 2.5 *Let μ_n be the law of $(X_1, \cdots, X_n)$. Then for any probability measure ν on E^n,*

$$W_1(\nu, \mu_n) \leq (1 + 2S_\phi)\sqrt{\frac{n}{2} \cdot h(\nu; \mu_n)}. \tag{2.9}$$

Here

$$W_1(\nu; \mu_n) := \inf \iint d_H(x, y) d\pi(x, y)$$

where the infimum is taken over all probability measures π on $E^n \times E^n$ with marginals ν and μ_n, is the Wasserstein distance between ν and μ_n; and

$$h(\nu, \mu_n) := \begin{cases} \int \log \frac{d\nu}{d\mu_n} d\nu & \text{if } \nu \ll \mu_n, \\ +\infty & \text{otherwise} \end{cases}$$

is the relative entropy (or Kullback information) of ν w.r.t. μ_n.

Notice that when $X_1, \cdots, X_n$ are independent, $S_\phi = 0$ and inequality (2.9) is a consequence of the Pinsky inequality together with the tensorization technique. Inequality (2.9) was proved at first by Marton [12] (1996) for Doeblin recurrent Markov chains, and next extended by Samson [15] (2000) for general ϕ-mixing sequences. But the condition in [15] is $\sum_k \sqrt{\phi_k} < +\infty$, stronger than the condition here.

See Ledoux [9] for a systematic treatment (and references) and application of such a transportation inequality to concentration of measure, and H. Djellout, A. Guillin and L. Wu [7] (2002) for some further extensions of Rio's result above.

2.4 Proofs of the main results

2.4.1 Proof of Theorem 2.3

Let

$$g(x_1, \cdots, x_n) := \int_{\mathbb{R}^d} \left| \frac{1}{n} \sum_{j=1}^{n} K_h(u - x_i) - f(u) \right| du, \ \forall x = (x_1, \cdots, x_n) \in E^n.$$

Then $D_n^* = g(X_1, \cdots, X_n)$. Fix $i = 1, \cdots, n$. For all $x, y \in E^n$ such that $x_j = y_j$ for all j except $j = i$, we have

$$|g(x) - g(y)| \leq \frac{1}{n} \int_{\mathbb{R}^d} |K_h(u - x_i) - K_h(u - y_i)| \, du \leq \frac{2}{n}.$$

In other words g verifies (2.6) with $L = 2/n$. Thus applying (2.8) to g and $-g$, we get (2.5).

2.4.2 Proof of Theorem 2.2

For the convenience of the reader we recall two well known lemmas in Analysis:

Lemma 2.6 *(L^1 version of Bochner's theorem) Let K be a nonnegative Borel function on $\mathbb{R}^d$ with $\int K(x)dx = 1$. Then $\lim_{h \to 0+} \int |K_h * f(x) - f(x)|dx = 0$, where $K_h(x) = h^{-d} K(x/h)$.*

See L.Devroye [5] (Lemma 1, p.897).

Lemma 2.7 *(Lebesgue density theorem)If f is a density on $\mathbb{R}^d$ and B is a compact set of $\mathbb{R}^d$ with $\lambda(B) > 0$ where λ is the Lebesgue measure, then*

$$\lim_{h \to 0} \lambda^{-1}(hB) \int_{x+hB} f(y)dy = f(x), \quad \textit{for almost all} \quad x.$$

See L.Devroye [5] (Lemma 2, p.898).

We now go to the

Proof. [Proof of Theorem 2.2]. The proof is divided into three steps, where the first two steps are close to that of [28].

Step 1 By Lemma 2.6, it suffices to show that $\int |f_n^*(x) - K_h * f(x)|dx \to 0$ exponentially as $n \to \infty$. Note that

$$f_n^*(x) = K_h * L_n = h^{-d} \int K(\frac{x-y}{h})dL_n(y).$$

Given $\varepsilon > 0$, we can find finite positive constants M, L, m, $a_1, \cdots, a_m$ and disjoint finite rectangles A_1, $\cdots$, A_m in $\mathbb{R}^d$ of form $\prod_{i=1}^d [x_i, x_i + a_i)$ such that the function

$$K^{(\varepsilon)}(x) = \sum_{j=1}^m a_j I_{A_j}(x)$$

satisfies: $K^{(\varepsilon)} \le M, K^{(\varepsilon)} = 0$ outside $[-L, L]^d$, and $\int |K(x) - K^{(\varepsilon)}|dx < \varepsilon$. Define

$$f_n^{(\varepsilon),*} := K_h^{(\varepsilon)} * L_n.$$

Then

$$\begin{aligned}
\int |f_n^*(x) - K_h * f(x)|dx \le & \int |f_n^*(x) - f_n^{(\varepsilon),*}(x)|dx \\
& + \int |f_n^{(\varepsilon),*}(x) - K_h^{(\varepsilon)} * f(x)|dx + \int |K_h^{(\varepsilon)} * f - K_h * f(x)|dx \\
\le & \int h^{-d} \int |K^{\varepsilon}(\frac{x-y}{h}) - K(\frac{x-y}{h})|L_n(dy)dx \\
& + \int |f_n^{(\varepsilon),*}(x) - K_h^{(\varepsilon)} * f(x)|dx \\
& + \int h^{-d} \int |K^{(\varepsilon)}(\frac{x-y}{h}) - K(\frac{x-y}{h})|f(y)dydx \\
\le & 2\varepsilon + \int |f_n^{(\varepsilon),*}(x) - K_h^{(\varepsilon)} * f(x)|dx.
\end{aligned} \tag{2.10}$$

Noting that $d\mu = f dx$, then

$$\begin{aligned}\int |f_n^{(\varepsilon),*}(x) - K_h^{(\varepsilon)} * f(x)|dx &\leq \sum_{j=1}^m |a_j| \int |h^{-d} \int_{x+hA_j} f(y)dy - h^{-d} \int_{x+hA_j} L_n(dy)|dx \\ &\leq Mh^{-d} \sum_{j=1}^m \int |\mu(x+hA_j) - L_n(x+hA_j)|dx.\end{aligned}$$

Consequently for Theorem 2.2, it is enough to prove that for any finite rectangle $A := \prod_{i=1}^d [x_i, x_i + a_i)$ of $\mathbb{R}^d$,

$$h^{-d} \int |L_n(x+hA) - \mu(x+hA)|dx \to 0 \quad \text{exponentially as} \quad n \to \infty. \tag{2.11}$$

Step 2. Fix such a rectangle $A := \prod_{i=1}^d [x_i, x_i + a_i)$, and let $\varepsilon > 0$ be arbitrary. Consider the partition of $\mathbb{R}^d$ into sets B that are d-fold products of intervals of the form $[\frac{(i-1)h}{N}, \frac{ih}{N})$, where i is an integer, and N is a fixed integer to be chosen later. Call the partition Ψ.

Let N be such that $\min_i a_i \geq \frac{2}{N}$, $A^* = \Pi_{i=1}^d [x_i + \frac{1}{N}, x_i + a_i - \frac{1}{N})$. Define

$$C_x = x + hA - \bigcup_{B \in \Psi, B \subseteq x+hA} B \subseteq x + h(A \backslash A^*).$$

Clearly,

$$\begin{aligned}&\int |\mu(x+hA) - L_n(x+A)|dx \\ &\leq \int \sum_{B \in \Psi, B \subseteq x+hA} |\mu(B) - L_n(B)|dx + \int \{\mu(C_x) + L_n(C_x)\}dx\end{aligned} \tag{2.12}$$

Using the fact that for any set C, and any probability measure ν on $\mathbb{R}^d$,

$$\int \nu(x+hC)dx = \lambda(hC),$$

where λ is the Lebesgue measure, the last term in (2.12) is bounded from above by

$$\begin{aligned}2\lambda(h(A\backslash A^*)) = 2h^d \lambda(A \backslash A^*) &= 2h^d \left(\prod_{i=1}^d a_i - \prod_{i=1}^d \left(a_i - \frac{2}{N}\right)\right) \\ &= 2h^d \lambda(A) \left(1 - \prod_{i=1}^d \left(1 - \frac{2}{Na_i}\right)\right) \\ &\leq \varepsilon h^d\end{aligned}$$

once if N verifies

$$\min_i a_i \geq \frac{2}{N}, \text{ and } 2\lambda(A)\left(1-\prod_{i=1}^{d}\left(1-\frac{2}{Na_i}\right)\right) \leq \varepsilon.$$

Fix such N which is independent of n.

For any finite constant $R > 0$, letting $S_{OR} := \{x \in \mathbb{R}^d;\ |x| \leq R\}$, we can bound the first term in (2.12) from above by

$$\sum_{B\in\Psi, B\cap S_{OR}\neq\phi} |L_n(B)-\mu(B)| \int_{B\subseteq x+hA} dx + \int_{B\subseteq x+hA} dx\{L_n(S_{OR}^c)-\mu(S_{OR}^c)+2\mu(S_{OR}^c)\}.$$

Here $(\cdot)^c$ denotes the complement of a set. Clearly, $h^{-d}\int_{B\subseteq x+hA} dx \leq \lambda(A)$, and $\mu(S_{OR}^c) < \varepsilon$ by sufficiently large R.

By Lemma 2.4,

$$\mathbb{P}\{L_n(S_{OR}^c) - \mu(S_{OR}^c) > \delta\} \leq \exp\left(-\frac{2n\delta^2}{(1+2S_\phi)^2}\right),\ \forall\delta > 0.$$

Consequently for (2.11) it remains to establish

$$\sum_{B\in\Psi, B\cap S_{OR}\neq\phi} |L_n(B) - \mu(B)| \to 0, \text{ exponentially.} \tag{2.13}$$

Step 3. Our proof of the key estimate (2.13) is very different from that in [5] and it is the main new point here. Set

$$\tilde{\Psi} = \{B; B \in \Psi, B \cap S_{OR} \neq \emptyset\},\ C := \left(\bigcup_{B\in\tilde{\Psi}} B\right)^c$$

$$\mathcal{B}(\tilde{\Psi}) = \sigma\{B; B \in \tilde{\Psi}\}.$$

Regarding L_n and μ as probability measures on $\mathcal{B}(\tilde{\Psi})$, and denoting the total variation of $L_n - \mu$ on $\mathcal{B}(\tilde{\Psi})$ by $\|L_n - \mu\|_{\mathcal{B}(\tilde{\Psi})}$, we have

$$\sum_{B\in\Psi, B\cap S_{OR}\neq\emptyset} |L_n(B) - \mu(B)| \leq \|L_n - \mu\|_{\mathcal{B}(\tilde{\Psi})} = 2\max_{\tilde{B}\in\mathcal{B}(\tilde{\Psi})} |L_n(\tilde{B}) - \mu(\tilde{B})|.$$

Therefore,

$$\begin{aligned}\mathbb{P}\left(\sum_{B\in\Psi, B\cap S_O R\neq\emptyset} |L_n(B) - \mu(B)| > \varepsilon\right) &\leq \mathbb{P}\left(\max_{\tilde{B}\in\mathcal{B}(\tilde{\Psi})} |L_n(\tilde{B}) - \mu(\tilde{B})| > \frac{\varepsilon}{2}\right)\\ &\leq \sum_{\tilde{B}\in\mathcal{B}(\tilde{\Psi})} \mathbb{P}\left(|L_n(\tilde{B}) - \mu(\tilde{B})| > \frac{\varepsilon}{2}\right)\end{aligned}$$

At first by Lemma 2.4, for each $\tilde{B} \in \mathcal{B}(\tilde{\Psi})$,

$$\mathbb{P}\left(|L_n(\tilde{B}) - \mu(\tilde{B})| > \frac{\varepsilon}{2}\right) \leq \mathbb{P}\left(L_n(\tilde{B}) > \mu(\tilde{B}) + \frac{\varepsilon}{2}\right) + \mathbb{P}\left(L_n(\tilde{B}) < \mu(\tilde{B}) - \frac{\varepsilon}{2}\right)$$
$$\leq 2\exp\left(-\frac{n\varepsilon^2}{2(1+2S_\phi)^2}\right).$$

Secondly, the number of elements $\#\tilde{\Psi}$ in $\tilde{\Psi}$ is not greater than $\left(\frac{2RN}{h}+2\right)^d = o(n)$ by **(H2)**, $\tilde{\mathcal{B}}(\phi)$ has $2^{\#\tilde{\Psi}} = 2^{o(n)}$ elements for n large enough. Consequently

$$\mathbb{P}\left(\sum_{B\in\Psi, B\cap S_O R\neq\emptyset} |L_n(B) - \mu(B)| > \varepsilon\right) \leq 2^{o(n)} 2\exp\left(-\frac{n\varepsilon^2}{2(1+2S_\phi)^2}\right)$$

where the desired (2.13) follows. This completes the proof of Theorem 2.2.
□

2.5 Concluding remarks

From the proof of Theorem 2.2 above, we see clearly that for proving the exponential convergence of f_n^* to the unknown density f in L^1, it is enough to show the key relation (2.13) together with the exponential convergence of $L_n(S_{OR}^c)$ to $\mu(S_{OR}^c)$. Thus by following the proof of (2.13), we see that Theorem 2.2 will remain valid once if we can prove the following exponential deviation inequality

$$\mathbb{P}(|L_n(A) - \mu(A)| > \delta) \leq C_1(\delta)e^{-C_2(\delta)n}, \ \forall A \subset \mathcal{B}(\mathbb{R}^d) \tag{2.14}$$

for some constants $C_1(\delta), C_2(\delta)$ depending only on δ (independent of n, A) and for any $\delta > 0$.

In this Note we have applied Rio's inequality which is stronger than (2.14). The reader certainly guess that (2.14) holds in a much wider situation than the uniform mixing case treated in this paper.

Bibliography

[1] Adams, T.M. and Nobel, A.B. (1998) On density estimation from ergodic processes. *Ann. Probab.*, **26**, 794-804.

[2] Bobkov, S.G. and Götze, F. (1999) Exponential Integrability and Transportation Cost Related to Logarithmic Sobolev Inequalities, *J. Funct. Anal.* **163**, 1-28.

[3] Bosq, D., Merlevède, F. and Peligrad, M. (1999) Asymptotic normality for density kernel estimators in discrete and continuous time. *J. Multivariate Anal.*, **68**, 78-95.

[4] Bryc, W. and Dembo, A. (1996) Large deviations and strong mixing, *Ann. Inst. H. Poincaré (Série Proba. et Stat.)*, *32*, 549-569.

[5] Devroye, L. (1983) The equivalence of weak, strong and complete convergence in L^1 for kernel density estimates, *Ann. Stat.*, **11**, 896-904.

[6] Devroye, L. (1988) The Kernel Estimate is Relatively Stable. *Prob. Theo. Rel. Field.*, **77**, 537-550.

[7] Djellout, H., Guillin, A. and Wu, L.M. (2004) Transportation cost-information inequalities for random dynamical systems and diffusions. *Ann. Probab.*, **32**, 2702-2732.

[8] Gao, F.G. (2001) Moderate Deviations and Large Deviations For the Kernel Density Estimators. Preprint.

[9] Ledoux, M. (1999) Concentration of measure and logarithmic sobolev inequalities, *S 'eminaire de Probab. XXXIII*, LNM, **1709**, 120-216.

[10] Lei, L.Z., Wu, L.M. and Xie, B. (2003) Large deviations and deviation inequality for kernel density estimator in $L_1(R^d)$-distance. In *Development of Modern Statistics and Related Topics*, Ser. Biostat. 1., 89-97, *World Sci. Publishing, River Edge, NJ.*

[11] Louani, D. (2000) Large deviations for the L_1-distance in kernel density estimation. *J. Statist. Plann. Inference.*, 177-182.

[12] Marton, K. (1996) A measure concentration inequality for contracting Markov chains.*Geom. Funct. Anal.*, **7**. 556-571.

[13] Peligrad, M. (1992) Properties of uniform consistency of the kernel estimators of density and of regression functions under dependence assumption. *Stoch. and Stoch. Reports*, **40**, 147-168.

[14] Rio, E. (2000) Inégalités de Hoeffding pour les fonctions lipchitziennes de suites dépendantes. *CRAS, Paris, t.330, Série I*, 905-908.

[15] Samson. P.M. (2000) Concentration of measure inequalites for Markov chains and ϕ-mixing processes. *Ann. Prob.*, **28**, 416-461.

Chapter 3

Large deviations of kernel density estimator in $L^1(\mathbb{R}^d)$ for uniformly ergodic Markov processes

(Published in: *Stochastic Processes and their Applications*)

This article is a joint work with Professor Liming Wu.

3.1 Introduction

Let $\{X_n; n \geq 0\}$ be a Doeblin recurrent Markov chain valued in a Borel measurable subset E of $\mathbb{R}^d$, defined on the probability space $(\Omega, (\mathcal{F}_n)_{n\in\mathbb{N}}, \mathcal{F}, (\mathbb{P}_x)_{x\in E})$, with (unknown) transition kernel $P(x, dy)$. Moreover, we assume that the unique invariant measure μ of P is absolutely continuous, i.e., $\mu(dx) = f(x)dx$ where the density f is unknown.

Let K be a measurable function such that

$$K \geq 0, \int_{\mathbb{R}^d} K(x)dx = 1, \tag{3.1}$$

and set $K_h(x) = \frac{1}{h^d}K(\frac{x}{h})$. Given the observed sample $\{X_0, \cdots, X_{n-1}\}$, we consider the empirical measure $L_n = \frac{1}{n}\sum_{i=0}^{n-1}\delta_{X_i}$ and define the kernel density estimator of the unknown f as usually as:

$$f_n^*(x) = K_{h_n} * dL_n(x) = \frac{1}{n}\sum_{i=0}^{n-1}\frac{1}{h_n^d}K\left(\frac{x-X_i}{h_n}\right), \quad x \in \mathbb{R}^d \tag{3.2}$$

where $\{h_n, n \geq 0\}$ is a sequence of positive numbers (bandwidth) satisfying

$$h_n \to 0, \quad nh_n^d \to +\infty \quad as \quad n \to \infty. \tag{3.3}$$

A natural distance of f_n^* from the unknown f is its $L^1(\mathbb{R}^d) := L^1(\mathbb{R}^d, dx)$ distance below,

$$D_n^* = \int_{\mathbb{R}^d} |f_n^*(x) - f(x)|dx. \tag{3.4}$$

The limit behavior of f_n^* in $L^1(\mathbb{R}^d)$ is a subject of current study.

In the *i.i.d.* case, Devroye [6] proved that all types of $L^1(\mathbb{R}^d)$-consistency are equivalent to the condition(3.3) on the bandwidth (h_n). Csörgö and Horváth [3] and Horváth [11] investigated the asymptotic normality of D_n^*. Louani [16] established the large deviation principle (LDP in short) of D_n^*. Gao [8] obtained the LDP and the moderate deviation principle of f_n^* in $L^\infty(\mathbb{R}^d)$. And recently Lei, Wu and Xie [14] prove the weak LDP of f_n^* in $L^1(\mathbb{R}^d)$, and show that the corresponding LDP is false. More recently Gao [9] obtains the moderate deviation principle of f_n^* in $L^1(\mathbb{R}^d)$ and the law of the iterated logarithm for D_n^*. Giné, Mason and Zaitsev [10] establish a functional central limit theorem and a Glivenko-Cantelli theorem.

How to extend those results from the i.i.d. case to Markov processes (or dependent case) is a very natural and important question. In fact, numerous practical models from economic time series or biologies are Markov process (cf. [2]), for which it is very important to estimate the asymptotic equilibrium measure $\mu(dx) = f(x)dx$. Known works in the dependent case are concentrated on the consistency of f_n^* and its asymptotic normality,see Peligrad [18] , Bosq, Merlevède and Peligrad [1] and the references therein. But little is known about the large deviations of f_n^* and D_n^* in the dependent case.
In a recent work [15], as a first step towards the large deviations of f_n^*, we prove the exponential convergence of f_n^* to f for a ϕ-mixing sequence (X_n). In this paper which is a sequel to [15], we investigate the large deviations of f_n^* in $L^1(\mathbb{R}^d)$ and of D_n^* in the framework of uniformly ergodic Markov chains (see **(H)** below).

Large deviation of occupation measures L_n for Markov processes is a traditional subject in probability, initiated by Donsker and Varadhan [7]. The rate function is the Donsker-Varadhan level-2 entropy given by

$$J(\nu) := \sup\left\{\int \log \frac{u}{Pu} d\nu;\ 1 \leq u \in b\mathcal{B}(E)\right\},\ \forall \nu \in M_1(E) \tag{3.5}$$

where $b\mathcal{B}(E)$ is the space of real bounded functions measurable w.r.t. the Borel σ-field $\mathcal{B}(E)$ of E, and $M_1(E)$ denotes the space of all probability measures on E.

Deuschel and Stroock [5] (Thm 4.1.14) obtained the LDP of L_n w.r.t. the τ-topology (i.e., the weakest topology on $M_1(E)$ such that $\nu \to \nu(f) := \int_E f(x)d\nu(x)$ is continuous for all $f \in b\mathcal{B}(E)$), under the following

(H) (*uniform ergodicity*) there exists $1 \le l \le N \in \mathbb{N}$ and $M \ge 1$ such that

$$P^l(x, A) \le M \frac{P(y, A) + \cdots + P^N(y, A)}{N}, \ \forall x, y \in E, A \in \mathcal{B}(E).$$

A lot of significant progresses have been made after, see [4], [23], [13] and the references therein.

This paper is organized as follows. The main results such as the weak*-LDP of f_n^* on $L^1(\mathbb{R}^d)$, the large deviation estimation for $\mathbb{P}_x(D_n^* > \delta)$ and the asymptotic efficiency of the estimator f_n^* in the Bahadur sense etc are presented in the next section. Those results are, as far as we know, obtained for the first time in the dependent case. In Section 3, we prepare several lemmas. We give the proofs of the main results in the last part: Section 4-7.

3.2 Main results

Throughout this paper, we adopt the following notations. $L^p(\mathbb{R}^d) := L^p(\mathbb{R}^d, dx)$, $L^p(\mu) := L^p(E, \mu)$; $\|f\|_1 = \|f\|_{L^1(\mathbb{R}^d, dx)}$. We denote by $b\mathcal{B}$ (resp. $b\mathcal{B}(E)$) the space of all real bounded and Borel $\mathcal{B}$-measurable functions on $\mathbb{R}^d$ (resp. E) equipped with the sup norm $\|V\| = \sup_x |V(x)|$. We write $\nu(V) = \langle V \rangle_\nu := \int_E V(x)d\nu(x)$. Without loss of generality, we assume that $(X_n)_{n \ge 0}$ is the system of coordinates on $\Omega := E^{\mathbb{N}}$ and $\mathbb{P}_x$ is the law of the Markov chain with the transition kernel P and the starting point $x \in E$. Set $\mathbb{P}_\nu(\cdot) := \int_E \mathbb{P}_x(\cdot)d\nu(x)$ and $\mathbb{E}^\nu(\cdot) = \int_\Omega \cdot \, d\mathbb{P}_\nu$. Let $(\theta\omega)_n := \omega_{n+1} (n \in \mathbb{N})$ be the shift on Ω.

When the bandwidth $h_n \to 0$, $f_n^* dx$ is "close" to L_n in the τ-topology, so we may hope that $f_n^* dx$ satisfies the same LDP as L_n. This intuition is true :

Theorem 3.1 *Assume* **(H)** *and* $h_n \to 0$ *(without (3.3)). Then* $\mathbb{P}_x(f_n^* \in \cdot)$ *satisfies, uniformly for the initial points* $x \in E$*, the LDP in* $L^1(\mathbb{R}^d)$ *w.r.t. the weak topology* $\sigma(L^1, L^\infty)$*, with the rate function given by*

$$J(g) := \begin{cases} J(gdx), & \text{if } g \in \mathcal{P}(E) \\ +\infty, & \text{if } g \in L^1(\mathbb{R}^d) \backslash \mathcal{P}(E). \end{cases} \tag{3.6}$$

Here $J(\cdot)$ *is the Donsker-Varadhan level-2 entropy given in (3.5),* $\mathcal{P}(E)$ *is the set of all probability density functions on* $\mathbb{R}^d$ *with support in* E*, i.e., those* $g \in L^1(\mathbb{R}^d)$ *such that* $g \ge 0$ *on* $\mathbb{R}^d$*,* $g = 0$*, a.e. on* $E^c := \mathbb{R}^d \backslash E$ *and* $\int_{\mathbb{R}^d} g dx = 1$*.*

More precisely, J is inf-compact on $(L^1(\mathbb{R}^d), \sigma(L^1, L^\infty))$, and for any measurable subset A of $L^1(\mathbb{R}^d)$,

$$
\begin{aligned}
- \inf_{g \in \overset{\circ}{A}{}^\sigma} J(g) &\le \liminf_{n\to\infty} \frac{1}{n} \log \inf_{x\in E} \mathbb{P}_x(f_n^* \in A) \\
&\le \limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in E} \mathbb{P}_x(f_n^* \in A) \le - \inf_{g\in \bar{A}^\sigma} J(g)
\end{aligned}
$$

where $\overset{\circ}{A}{}^\sigma, \bar{A}^\sigma$ denote respectively the interior and the closure of A w.r.t. the weak topology $\sigma(L^1, L^\infty)$.

The LDP w.r.t. the weak topology on $L^1(\mathbb{R}^d)$ above is of the same type as the classical results for L_n w.r.t. the τ-topology. But it is too weak in the sense that it does not entail the consistency, i.e., $D_n^* \to 0$ in probability. For statistical issues, the main objects to be studied are

(i) $\mathbb{P}_x(\|f_n^* - g\|_1 < \delta)$ where $g \in \mathcal{P}(E)$ is fixed, which is important in the hypothesis testing: $H_0 :\ d\mu(x) = f(x)dx$ against $H_1 :\ d\mu(x) = g(x)dx$; or

(ii) $\mathbb{P}_x(D_n^* > \delta)$, whose statistical importance is obvious.

Unfortunately Theorem 3.1 can not be applied for them, since $\{\tilde{g} \in L^1(\mathbb{R}^d);\ \|\tilde{g} - g\|_1 < \delta\}$ is not open in $\sigma(L^1, L^\infty)$ and $\{\tilde{g} \in L^1(\mathbb{R}^d);\ \|\tilde{g} - f\|_1 \ge \delta\}$ is not closed in $\sigma(L^1, L^\infty)$. They are objects of

Theorem 3.2 *Assume* **(H)** *and (3.3). Then $\mathbb{P}_x(f_n^* \in \cdot)$ satisfies, uniformly for initial state $x \in E$, the weak*-LDP on $(L^1(\mathbb{R}^d), \|\cdot\|_1)$ with the rate function $J(g)$ given by (3.6), i.e., for any $g \in L^1(\mathbb{R}^d)$,*

$$
\begin{aligned}
&\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \inf_{x\in E} \mathbb{P}_x(\|f_n^* - g\|_1 < \delta) \\
&= \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in E} \mathbb{P}_x(\|f_n^* - g\|_1 < \delta) = -J(g).
\end{aligned}
\tag{3.7}
$$

Notice that the corresponding (good) LDP is in general not true, because even in the i.i.d. case, $J(g) = J^{iid}(g) = \int g(x) \log \frac{g(x)}{f(x)} dx$ (for $g \in \mathcal{P}(E)$ and $gdx \ll fdx$) is not inf-compact on $(L^1(\mathbb{R}^d), \|\cdot\|_1)$ (as noted in [14]).

Theorem 3.3 *Assume* **(H)** *and (3.3). Then*

(a) For any $\delta > 0$,

$$
\begin{aligned}
-I(\delta) &\le \liminf_{n\to\infty} \frac{1}{n} \log \inf_{x\in E} \mathbb{P}_x(\|f_n^* - f\|_1 > \delta) \\
&\le \limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in E} \mathbb{P}_x(\|f_n^* - f\|_1 > \delta) \le -I(\delta-)
\end{aligned}
\tag{3.8}
$$

where

$$I(\delta) = \inf\{J(g) | g \in \mathcal{P}(E), \|g - f\|_1 > \delta\}. \tag{3.9}$$

(b) We have for any $\delta > 0$,

$$I(\delta) \geq \frac{1}{l}\left(I^{iid}(\delta) - \log M\right) \tag{3.10}$$

where l, M are given in **(H)**, *and $I^{iid}(\delta)$ is the rate function of the LDP of $\|f_n^* - f\|_1$ in the case where (X_n) are i.i.d. of common law μ (see (3.14) below).*

(c) Besides **(H)**, *assume that P is aperiodic. Then we also have*

$$I(\delta) \geq \frac{\delta^2}{2(1+S)^2}, \ \forall \delta > 0 \tag{3.11}$$

where $S := \sum_{k=1}^{\infty} \sup_{x,y \in E} \|P^k(x, \cdot) - P^k(y, \cdot)\|_{TV}$ (here $\|\cdot\|_{TV}$ denotes the total variation) is finite.

Remarks (2.i) Parts (b) and (c) of Theorem 3.3 are served for δ large or small, respectively. By the contraction principle and the LDP of L_n under **(H)** in [5] (Thm 4.1.14), for each $V \in b\mathcal{B}(E)$, $L_n(V) - \mu(V)$ satisfies the LDP with the inf-compact rate function given by

$$J_V(r) = \inf\{J(\nu);\ \nu(V) = \mu(V) + r\},\ \forall r \in \mathbb{R}. \tag{3.12}$$

Since $J_V(0) = 0$ and J_V is convex with values in $[0, +\infty]$, J_V is nondecreasing and left continuous on $[0, +\infty)$. Consequently using $\|\nu - \mu\|_{TV} = \sup_{\|V\| \leq 1}[\nu(V) - \mu(V)] = 2\sup_{A \in \mathcal{B}} |\nu(A) - \mu(A)|$ (for two probability measures μ, ν), we can identify $I(\delta)$ given in (3.9) as

$$\begin{aligned} I(\delta) &= \inf\{J(\nu) | \sup_{\|V\| \leq 1} [\nu(V) - \mu(V)] > \delta\} \\ &= \inf_{\|V\| \leq 1} \inf_{r > \delta} J_V(r) = \inf_{\|V\| \leq 1} J_V(\delta+) \\ &= \inf\{J(\nu) | \sup_{A \in \mathcal{B}(E)} [(\nu(A) - \mu(A)] > \delta/2\} = \inf_{A \in \mathcal{B}(E)} J_A(\delta/2+) \end{aligned} \tag{3.13}$$

where $J_A = J_{1_A}$. In the i.i.d. case, the last expression in (3.13) above coincides exactly with the rate function of the LDP for D_n^* found by Louani [16]. Indeed, when $\mu(A) = a \in (0,1)$, then for any $\delta > 0$, $J_A^{iid}(\delta/2)$ is given by

$$\Gamma_a^+(\delta) = \begin{cases} (a + \frac{\delta}{2}) \log(1 + \frac{\delta}{2a}) + (1 - a - \frac{\delta}{2}) \log(1 - \frac{\delta}{2(1-a)}) & \text{if } 0 < \delta < 2 - 2a \\ +\infty, & \text{otherwise} \end{cases}$$

(then $J_A^{iid}(\delta/2) = J_A^{iid}(\delta/2+)$) and

$$I^{iid}(\delta) := \inf_{a\in(0,1)} \Gamma_a^+(\delta) = \inf_A J_A^{iid}(\delta/2) \tag{3.14}$$

which is I^{iid} in [16].

Remarks (2.ii) If I were increasing on $(0, a)$ where $a := \sup\{r > 0; I(r) < +\infty\}$, then we can prove in fact the LDP of D_n^* in $\mathbb{R}^+$ with the rate function $\delta \to I(\delta-)$, from (3.8).

In the results above, we have the large deviation estimates of the estimator f_n^*, useful in statistics. We now show that f_n^* is asymptotically optimal in the Bahadur sense. Let Θ be the set of unknown data (P, μ) verifying **(H)** and $\mu(dx) \ll dx$. Given a subset $\mathcal{D}$ of the unit ball in $b\mathcal{B}$, we say that an estimator $T_n(\cdot) := T_n(\cdot; X_0, \cdots, X_{n-1}) \in L^1(\mathbb{R}^d)$ is an asymptotically $\sigma(L^1, \mathcal{D})$-consistent estimator of the density f, if $\forall V \in \mathcal{D}$,

$$\int_{\mathbb{R}^d} T_n(x)V(x)dx \to \int_{\mathbb{R}^d} f(x)V(x)dx$$

in probability measure $\mathbb{P}_\mu$. From the results above, we shall derive:

Theorem 3.4 *Given $(P, \mu) \in \Theta$, let $((X_n), (\mathbb{P}_x)_{x\in E})$ be the associated Markov process.*

(a) **(Bahadur type lower bound)** *Assume that $\mathcal{D}$ is dense in the unit ball of $L^\infty(\mathbb{R}^d)$ w.r.t. the weak* topology $\sigma(L^\infty, L^1)$. Then for any $\sigma(L^1, \mathcal{D})$-asymptotically consistent estimator T_n of the density f,*

$$\begin{aligned} &\liminf_{r\to 0+} \frac{1}{r^2} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\|T_n - f\|_1 > r) \\ &\geq -\frac{1}{2\sup_{\|V\|\leq 1} \sigma^2(V)} = -\frac{1}{8\sup_{A\in\mathcal{B}} \sigma^2(1_A)} \end{aligned} \tag{3.15}$$

where

$$\sigma^2(V) := Var_\mu(V) + 2\sum_{k=1}^\infty \langle V - \mu(V), P^k V\rangle_\mu.$$

If moreover $\|T_n - T_n \circ \theta^N\|_1 \leq \delta_n \to 0$, then (3.15) still holds with $\mathbb{P}_\mu$ substituted by $\inf_{x\in E} \mathbb{P}_x$.

(b) **(Asymptotic efficiency of f_n^* in the Bahadur sense)** *If h_n verifies (3.3), then*

$$\begin{aligned}&\liminf_{r\to 0+}\frac{1}{r^2}\lim_{n\to\infty}\frac{1}{n}\log\inf_{x\in E}\mathbb{P}_x(\|f_n^*-f\|_1>r)\\&=\limsup_{r\to 0+}\frac{1}{r^2}\lim_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{P}_x(\|f_n^*-f\|_1>r)\\&=-\frac{1}{2\sup_{\|V\|\le 1}\sigma^2(V)}=-\frac{1}{8\sup_{A\in\mathcal{B}}\sigma^2(1_A)}.\end{aligned}\tag{3.16}$$

Thus f_n^* is an asymptotically efficient estimator of f in the Bahadur sense. And $1/\sigma^2(V)$ can be interpreted as the Fisher information at the direction V of our statistical model Θ.

All the results above except perhaps Theorem 3.4(a) are, as far as we know, new in the dependent case.

Remarks (2.iii) In comparison with the i.i.d. case, the new object in the Markov chain case is the transition kernel density $p(x,y) := P(x,dy)/dy$. For its estimation or more precisely $F(x,y) := f(x)p(x,y)$, no more effort is required due to the subtleness of our assumption **(H)**. Indeed, consider the Markov chain $Y_n := (X_n, X_{n+1})$ with values in E^2, whose transition kernel still verifies **(H)** and whose unique invariant measure is $F(x,y)dxdy$. The Donsker-Varadhan level-2 entropy for this new Markov chain possesses an explicit expression ([5]):

$$J^{(2)}(Q) := \begin{cases}\iint_{E\times E} Q(dx,dy)\log\frac{Q(x,dy)}{P(x,dy)}, & \text{if } Q\in M_1^s(E^2),\ Q(x,\cdot)\ll P(x,\cdot)\\ +\infty & \text{otherwise}\end{cases}\tag{3.17}$$

where $Q\in M_1^s(E^2)$ iff $Q\in M_1(E^2)$ and $Q(A\times E)=Q(E\times A), \forall A\in\mathcal{B}(E)$, and $Q(x,dy)$ is the regular conditional distribution of the second coordinate X_1 knowing the first $X_0=x$. Consider the kernel density estimator

$$F_n^*(x,y) := \frac{1}{n}\sum_{k=0}^{n-1} K_{h_n}(x-X_k)\cdot K_{h_n}(y-X_{k+1}).$$

Hence the previous results apply for F_n^* if the condition (3.3) is substituted by $h_n\to 0$ and $nh_n^{2d}\to+\infty$.

3.3 Several lemmas

For every $V\in b\mathcal{B}(E)$, put $P^V(x,dy) := e^{V(x)}P(x,dy)$. We have the Feynman-Kac formula

$$(P^V)^n f(x) = \mathbb{E}^x f(X_n)\exp\sum_{k=0}^{n-1}V(X_k).$$

Let $\|(P^V)^n\| := \sup_{\|f\|\le 1}\|(P^V)^n f\| = \|(P^V)^n 1\|$ be the norm of P^V acting on $b\mathcal{B}(E)$. Consider the uniform Cramer functional ([5])

$$\Lambda(V) = \lim_{n\to\infty}\frac{1}{n}\log\|(P^V)^n\| = \lim_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{E}^x\exp\left(\sum_{k=0}^{n-1}V(X_k)\right),$$

then $e^{\Lambda(V)}$ is the spectral radius of P^V on $b\mathcal{B}(E)$. It is well known ([5]) that

$$J(\nu) = \sup\{\nu(V)-\Lambda(V)|V\in b\mathcal{B}(E)\},\ \forall \nu\in M_1(E). \tag{3.18}$$

By the LDP of L_n in [5] and the Laplace principle due to Varadhan, $\forall V\in b\mathcal{B}(E)$,

$$\Lambda(V) = \sup\{\nu(V)-J(\nu)|\nu\in M_1(E)\} = \sup\left\{\int gV d\mu - J(g)|g\in\mathcal{P}(E)\right\}, \tag{3.19}$$

where the second equality follows from the fact that if $J(\nu)<+\infty$, then $\nu\ll\mu$ under **(H)** (see [22] (B.23)).

By **(H)**, $P^l(x,dy)\le M\mu(dy)$. Hence for each $V\in b\mathcal{B}(E)$,

$$\Lambda(V) = \lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}^\mu\exp\left(\sum_{k=0}^{n-1}V(X_k)\right). \tag{3.20}$$

Lemma 3.5 *For positive operator P^V defined as above, let $(P^V)^*$ be the dual operator of P^V w.r.t. μ. Then*

(a) There exist $\phi\in b\mathcal{B}(E)$, $\psi\in b\mathcal{B}(E)$ both strictly positive, such that

$$P^V\phi = e^{\Lambda(V)}\phi \quad over \quad E, \qquad (P^V)^*\psi = e^{\Lambda(V)}\psi,\ \mu-a.s.$$

and the following Harnack inequalities hold:

$$\frac{\phi(y)}{\phi(x)}\vee\frac{\psi(y)}{\psi(x)} \le \frac{M}{N}\cdot e^{2N\|V\|}\cdot\frac{\sum_{k=1}^N e^{k\Lambda(V)}}{e^{l\Lambda(V)}} \le Me^{3N\|V\|},\ \forall x,y\in E. \tag{3.21}$$

(b) Put

$$Q^V(x,dy) = \frac{\phi(y)}{e^{\Lambda(V)}\phi(x)}e^{V(x)}P(x,dy),$$

then Q^V is Doeblin recurrent, and $\nu_V := \phi\psi\mu$ is the unique invariant probability measure for Q^V.

Proof. (a) Under **(H)**, $P^l(x,dy) \le M\mu(dy)$ and then $P^N(x,dy) \le M\mu(dy)$. Thus $(P^V)^N$ is uniformly integrable in $L^\infty(\mu)$ in the terms of [23]. By Theorem 3.2 in [23], there exists some $0 \le \varphi \in L^\infty(\mu)$ such that $\mu(\varphi) > 0$ and

$$(P^V)^N \varphi = r^N \varphi, \ \mu - a.s.$$

where r is the spectral radius of P^V in $L^\infty(\mu)$. Since $(P^V)^N(x,dy) \le e^{N\|V\|} M\mu(dy)$, then letting $g := (P^V)^N\varphi$, we see that $(P^V)^N g = r^N g$ everywhere over E. By (3.20), $r = e^{\Lambda(V)}$. Finally setting

$$\phi(x) = \sum_{k=1}^{N} (P^V)^k g(x),$$

which is strictly positive by **(H)**, we have for all $x \in E$,

$$P^V \phi(x) = r\phi(x) = e^{\Lambda(V)}\phi(x), \ \forall x \in E.$$

Since for any x, y,

$$\frac{\phi(y)}{\phi(x)} = \frac{\left(P^V\right)^l \phi(y)}{\sum_{k=1}^N (P^V)^k \phi(x)} \cdot \frac{\sum_{k=1}^N e^{k\Lambda(V)}}{e^{l\Lambda(V)}},$$

using **(H)** and $-\|V\| \le V(x) \le \|V\|$, we get

$$\frac{\phi(y)}{\phi(x)} \le \frac{M}{N} \cdot e^{2N\|V\|} \cdot \frac{\sum_{k=1}^N e^{k\Lambda(V)}}{e^{l\Lambda(V)}}$$

where the desired Harnack inequality (3.21) for ϕ follows.

For the corresponding result about $(P^V)^*$, we choose a kernel $P^*(x,dy)$, which is the dual of P (w.r.t. μ) and also satisfies **(H)**. Applying the previous argument to $e^{V(y)}P^*(x,dy)$ which is the dual of P^V (w.r.t. μ), we get the existence of ψ and the Harnack inequality (3.21) for ψ .

(b) It is easy to verify that Q^V is a Markov kernel, and $\phi\psi\mu$ is an invariant measure of Q^V. As Q^V again satisfies **(H)** by part (a), it is Doeblin recurrent. Then $\phi\psi\mu$ is the unique invariant measure of Q^V. □

Lemma 3.6 *Under* **(H)**, *we have for every $V \in b\mathcal{B}(E)$ such that $\|V\| \le 1$, $\forall r > 0,\ n \ge 1$ so that $4N/n \le r$,*

$$\sup_{x\in E} \mathbb{P}_x \left(\frac{1}{n} \sum_{k=0}^{n-1} V(X_k) > \mu(V) + r \right) \le M \exp\left(-nJ_V \left(r - \frac{4N}{n} \right) \right), \qquad (3.22)$$

where $J_V(r)$ is the rate function governing the LDP of $L_n(V) - \mu(V)$, given in (3.12).

Notice that in the i.i.d. case, $M = N = 1$ and (3.22) is exactly the well known Cramer inequality. This lemma is basic to Theorem 3.3.

Proof. (following closely [5]) **1)** At first by Deuschel and Stroock [5] Lemma 4.1.4,

$$p_n(r) := \inf_{x\in E} \mathbb{P}_x \left(\frac{1}{n} \sum_{k=0}^{n-1} V(X_k) > \mu(V) + r \right)$$

is super-multiplicative, i.e., $p_{n+m} \geq p_n p_m,\ \forall n, m \in \mathbb{N}^*$. Thus

$$\frac{1}{n} \log p_n(r) \leq \sup_{m\geq 1} \frac{\log p_m(r)}{m} = \lim_{m\to\infty} \frac{\log p_m(r)}{m}.$$

But by the uniform LDP of $L_n(V) = \frac{1}{n}\sum_{k=0}^{n-1} V(X_k)$ in [5] and the increasingness of J_V on $\mathbb{R}^+$, we have $\lim\limits_{m\to\infty} \dfrac{\log p_m(r)}{m} \leq -J_V(r)$ for every $r \geq 0$. Thus

$$\inf_{x\in E} \mathbb{P}_x \left(\frac{1}{n} \sum_{k=0}^{n-1} V(X_k) > \mu(V) + r \right) \leq e^{-nJ_V(r)},\ \forall n \geq 1, r \geq 0. \tag{3.23}$$

2) For every $k = 1, \cdots, N$, since

$$|L_n(V) \circ \theta^k - L_n(V)| \leq \frac{2k}{n} \leq \frac{2N}{n},$$

letting $\varepsilon = \frac{2N}{n}$, we have for any $r \in \mathbb{R}$, $n \geq 1$ and $x \in E$,

$$\begin{aligned} f_{n,r}(x) := \mathbb{P}_x \left(L_n(V) > \mu(V) + r \right) &\leq \mathbb{P}_x \left(L_n(V) \circ \theta^k > \mu(V) + r - \varepsilon \right) \\ &= (P^k f_{n,r-\varepsilon})(x) \end{aligned}$$

and similarly

$$f_{n,r}(x) \geq \mathbb{P}_x \left(L_n(V) \circ \theta^k > \mu(V) + r + \varepsilon \right) = (P^k f_{n,r+\varepsilon})(x).$$

Thus using **(H)**, we obtain for any $x, y \in E$,

$$f_{n,r}(x) \leq (P^l f_{n,r-\varepsilon})(x) \leq M \frac{1}{N} \sum_{k=1}^{N} (P^k f_{n,r-\varepsilon})(y) \leq M f_{n,r-2\varepsilon}(y).$$

Hence the desired result follows by (3.23). □

The following result is technically crucial for all results in this paper.

Lemma 3.7 *(a) $\Lambda(V)$ is Gateaux-differentiable on $b\mathcal{B}(E)$.*

(b) If $V_n \to V$ in measure μ and $\sup_n \|V_n\| \le C$, then $\Lambda(V_n) \to \Lambda(V)$.

Proof. (a) Under **(H)**, $(P^V)^N$ is uniformly integrable in $L^\infty(\mu)$, then by [23], Proposition 2.1, $(P^V)^{2N}$ is compact in $L^\infty(\mu)$. Consequently by the perturbation theory of linear operators [12](Chap.VII, Theorem 1.8), the largest eigenvalue $e^{2N\Lambda(V)}$ of $(P^V)^{2N}$, is real-analytic, i.e., $\Lambda(V + t\tilde{V})$ is analytic on $t \in \mathbb{R}$ for any $V, \tilde{V} \in b\mathcal{B}$ fixed.

(b) At first $\liminf_{n\to\infty} \Lambda(V_n) \ge \Lambda(V)$ by (3.19). Notice that $e^{N\Lambda(V)}$ is the spectral radius of $(P^V)^N$ in $L^\infty(\mu)$. Now the inverse inequality $\limsup_{n\to\infty} \Lambda(V_n) \le \Lambda(V)$, follows by [23], Prop. 3.8. applied to $\pi_n := (P^{V_n})^N$. $\square$

Lemma 3.8 *(Gibbs type principle) Given a function $V \in b\mathcal{B}(E)$, a probability measure ν on E satisfies*

$$J(\nu) = \langle \nu, V\rangle - \Lambda(V)$$

iff $\nu = \nu_V := \phi\psi\mu$, where ϕ (resp. ψ) is the right (resp. left) eigenfunction of P^V associated with $e^{\Lambda(V)}$ given in Lemma 3.5(a) verifying $\mu(\phi\psi) = 1$.

Proof. Recall at first that

$$J(\nu) = \inf\{J^{(2)}(Q) | Q \in M_1^s(E^2), Q(A \times E) = \nu(A),\ \forall A \in \mathcal{B}(E)\} \tag{3.24}$$

where $J^{(2)}(Q)$ is given in (3.17) (cf. [7], [5]).

"$\Longleftarrow$" Let $\mathbb{Q}^V(dx, dy) = \nu_V(dx) Q^V(x, dy)$. By the definition (3.17), we have

$$\begin{aligned} J^{(2)}(\mathbb{Q}^V) &= \mathbb{E}^{\mathbb{Q}^V} \log \frac{Q^V(x,dy)}{P(x,dy)} = \mathbb{E}^{\mathbb{Q}^V} \log \frac{\phi(y)}{e^{\Lambda(V)}\phi(x)} \cdot e^{V(x)} \\ &= \int \log \frac{e^{V(x)}}{e^{\Lambda(V)}} d\nu_V(x) = \langle V, \nu_V\rangle - \Lambda(V) \end{aligned} \tag{3.25}$$

By (3.24), $J(\nu_V) \le \langle V, \nu_V\rangle - \Lambda(V)$ and the equality holds by (3.18).

"$\Longrightarrow$" It is well known from the convex analysis that

$$J(\nu) = \langle \nu, V\rangle - \Lambda(V) \iff \nu \in \partial\Lambda(V) \tag{3.26}$$

where $\partial\Lambda(V)$ denotes the set of sub-differentials of $\Lambda(\cdot)$ at V (which is contained in the topological dual space $(b\mathcal{B}(E))'$ to which $M_1(E)$ is embedded). Since $\nu_V \in \partial\Lambda(V)$ (by the sufficiency above) and $\Lambda(V)$ is Gateaux-differentiable on $b\mathcal{B}$ by Lemma 3.7, $\partial\Lambda(V)$ is the singleton $\{\nu_V\}$. $\square$

The following lemma is a main result in [15], which will be crucial in the proof of the lower bound in Theorem 3.2.

Lemma 3.9 ([15], Theorem 2.1) *Given a stationary sequence $(X_i)_{i\in\mathbb{N}}$ valued in E such that $\mu(dx) = \mathbb{P}(X_i \in dx) = f(x)dx$. Let $(\phi_k)_{k\geq 1}$ be the ϕ-mixing coefficient of $(X_i)_{i\in\mathbb{N}}$. Assume (3.3) and*

$$S_\phi := \sum_{k=1}^{\infty} \phi_k < +\infty. \tag{3.27}$$

Let D_n^* be given by (1.2). Then $D_n^* \to 0$ exponentially as $n \to \infty$, i.e.,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(D_n^* > \delta) < 0, \quad \forall \delta > 0.$$

Corollary 3.10 *If P is a Doeblin recurrent* ([17]) *Markov kernel on E with the unique invariant probability measure $d\mu(x) = f(x)dx$, then*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(D_n^* > \delta) < 0, \quad \forall \delta > 0.$$

Proof. If P is moreover aperiodic, then $S_\phi < +\infty$ (well known, see the proof of Theorem 3.3(c) in Section 6) and this corollary follows directly from Lemma 3.9. Now assume that P is of period $d > 1$. By the classical theory of Markov chains in [17], we have the following cyclic decomposition: $E = \mathcal{N} \bigcup E_1 \bigcup \cdots \bigcup E_d$ where $\mu(\mathcal{N}) = 0$ and

(i) $\mathcal{N}, E_1, \cdots, E_d$ are disjoint;

(ii) $P(x, E_{i+1}) = 1,\ \forall x \in E_i$ (here $E_{d+1} := E_1$);

(iii) there are $C > 0$ and $r \in (0, 1)$ such that

$$\sup_{x\in E_i} \|P^{nd}(x, \cdot) - \mu_i\|_{TV} \leq Cr^n,\ \forall n \geq 0, i = 1, \cdots, d$$

where $f_i = f1_{E_i}d$ and $\mu_i = d1_{E_i}\mu = f_i dx$. Let

$$f_{n,d}^*(x) := \frac{1}{n} \sum_{k=0}^{n-1} K_{h_n}(x - X_{dk}).$$

Since $P^d|_{E_i}$ is Doeblin recurrent and aperiodic on E_i by property (iii) above, we have by Lemma 3.9,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{\mu_i}(\|f_{n,d}^* \circ \theta^j - f_{i+j}\|_1 > \delta) < 0, \quad \forall \delta > 0$$

for all $i, j = 1, \cdots, d$ where $i + j := i + j(mod\ d)$. As $f^*_{nd}(x) = \frac{1}{d}\sum_{j=1}^d f^*_{n,d} \circ \theta^j$ and $f = \frac{1}{d}\sum_{j=1}^d f_{i+j}$, then we get for any $\delta > 0$ and $i = 1, \cdots, d$

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{nd} \log \mathbb{P}_{\mu_i}(\|f^*_{nd} - f\|_1 > \delta) \\
&\le \limsup_{n\to\infty} \frac{1}{nd} \log \sum_{j=1}^d \mathbb{P}_{\mu_i}(\|f^*_{n,d} \circ \theta^j - f_{i+j}\|_1 > \delta) < 0
\end{aligned}$$

where the desired result follows. □

Lemma 3.11 *Under* **(H)**, *we have:*

(a) for any $k \ge 1$, there exists some $\delta > 0$ such that

$$\sup_{|t|\le\delta} \sup_{\|V\|\le 1} \left| \frac{d^k}{dt^k} \Lambda(tV) \right| < +\infty$$

and for every $V \in b\mathcal{B}(E)$, $\Lambda''(tV)|_{t=0} = \sigma^2(V)$.

(b) the rate function J_V given in (3.12) satisfies

$$J_V(r) = \begin{cases} \sup_{t\in\mathbb{R}} (t[(r + \mu(V)] - \Lambda(tV)), & \forall r \in \mathbb{R}, \\ \sup_{t\ge 0} (t[(r + \mu(V)] - \Lambda(tV)), & \forall r \ge 0, \end{cases} \tag{3.28}$$

and J_V is strictly convex on $[J_V < +\infty]^0 = (a, b)$ where $a = \lim_{t\to-\infty} \frac{d}{dt}\Lambda(tV) - \mu(V)$ and $b = \lim_{t\to+\infty} \frac{d}{dt}\Lambda(tV) - \mu(V)$ (in particular J_V is strictly increasing and continuous in $[0, b)$); moreover

$$\lim_{r\to 0+} \frac{J_V(r)}{r^2} = \frac{1}{2\sigma^2(V)} \in (0, +\infty].$$

Proof. (a) We shall follow the approach in [21], in which it is assumed that 1 is the unique isolated eigenvalue $z \in \mathbb{C}$ of P in $b\mathcal{B}(E)$ such that $|z| = 1$. Under **(H)**, the last assumption is satisfied if P is aperiodic. Let us see how to bypass this assumption.

Under **(H)**, recall the cyclic decomposition $E = \mathcal{N} \bigcup \bigcup_{i=1}^d E_i$ in the proof of Corollary 3.10 above. Let us consider $P^d|_{E_i}$ which is Doeblin recurrent on E_i, aperiodic, with the unique invariant probability measure μ_i. Hence 1 is the unique isolated eigenvalue $z \in \mathbb{C}$ of $P^d|_{E_i}$ in $b\mathcal{B}(E_i)$ such that $|z| = 1$.

For each $V \in b\mathcal{B}(E)$, consider the following operator acting on $b\mathcal{B}(E_1)$:

$$R^V f(x) := \mathbb{E}^x f(X_d) e^{\sum_{k=0}^{d-1} V(X_k)} = (P^V)^d|_{E_1} f(x), \ \forall x \in E_1.$$

It is obvious that the spectral radius $r_{sp}(R^V)$ of R^V in $b\mathcal{B}(E_1)$ is not greater than $r_{sp}((P^V)^d) = e^{d\Lambda(V)}$. On the other hand, by the LDP in [5] for any initial measure and the fact that $(P^V)^d 1_{E_1^c} = 0$ on E_1, we have

$$\begin{aligned}
\log r_{sp}(R^V) &\geq \lim_{n\to\infty} \frac{1}{n} \log \mu_1[(P^V)^{nd} 1_{E_1}] \\
&= \lim_{n\to\infty} \frac{1}{n} \log \mu_1[(P^V)^{nd} 1] \\
&= d\Lambda(V).
\end{aligned}$$

Thus $r_{sp}(R^V) = e^{d\Lambda(V)}$.

As in [21], we will apply the analytical perturbation theory in Kato [12]. For each $z \in \mathbb{C}$, consider R^{zV} acting on the complexified space $b_{\mathbb{C}}\mathcal{B}(E_i)$, which is analytical in z in the sense of [12]. Then for any $\eta \in (0, 1/2)$ sufficiently small, there exists $\delta > 0$ and $C > 0$ such that for all $V \in b\mathcal{B}(E)$ with $\|V\| \leq 1$,

1) the eigenvalue $\lambda_{max}(R^{zV})$ of R^{zV} with the largest modulus is isolated in the spectrum of R^{zV} and $|\lambda_{max}(R^{zV}) - 1| \leq \eta$ for $|z| \leq 2\delta$;

2) for all $|z| \leq 2\delta$, the eigenprojection $E(z, V)$ of R^{zV} associated with $\lambda_{max}(R^{zV})$ is unidimensional and

$$\|E(z, V)1_{E_1} - 1_{E_1}\| < 1/2, \ \|(R^{zV})^n (I - E(z, V))\| \leq C(1 - 2\eta)^{nd}, \ \forall n;$$

3) $z \to \lambda_{max}(R^{zV})$ and $z \to E(z, V)f$ is analytic in z for $|z| \leq 2\delta$;

where properties 1) and 2) follow by [12], Chap.IV, Theorem 3.16, and the property 3) by [12], Chap.VII, Theorem 1.8..

Then $\Lambda(zV) := \frac{1}{d} \log \lambda_{max}(R^{zV})$ is analytic for $|z| \leq 2\delta$ and coincides with $\Lambda(tV)$ when $z = t \in [-2\delta, 2\delta] \subset \mathbb{R}$.

Let $\Lambda_n(zV) := \frac{1}{nd} \log \mathbb{E}^{\mu_1} \exp(\sum_{k=0}^{nd-1} zV(X_k)) = \frac{1}{nd} \log\langle 1, (R^{zV})^n 1\rangle_{\mu_1}$. By the properties 1) and 2) above, we have

$$\langle 1, (R^{zV})^n 1\rangle_{\mu_1} = e^{nd\Lambda(zV)} \langle 1, E(z, V)1\rangle_{\mu_1} + O((1 - 2\eta)^{nd})$$

where it follows that $\Lambda_n(zV) \to \Lambda(zV)$ uniformly over $z: \ |z| \le 2\delta$ and $V: \|V\| \le 1$. Thus by Cauchy's theorem and the property 3) above,

$$\sup_{\|V\|\le 1} \sup_{|z|\le\delta} \left| \frac{d^k}{dz^k}\Lambda(zV) \right| < +\infty,$$

$$\sup_{\|V\|\le 1} \sup_{|z|\le\delta} \left| \frac{d^k}{dz^k}\Lambda_n(zV) - \frac{d^k}{dz^k}\Lambda(zV) \right| \to 0.$$

Applying the above estimate to $k=2$ and notice that $\mathbb{E}^{\mu_i}\sum_{k=1}^d V(X_k) = d\mu(V)$,

$$\begin{aligned}\Lambda_n''(tV)|_{t=0} &= \frac{1}{nd}\mathbb{E}^{\mu_1}\left(\sum_{k=0}^{nd-1} V(X_k) - \mathbb{E}^{\mu_1}\sum_{k=0}^{nd-1} V(X_k)\right)^2 \\ &\to Var_{\mathbb{P}_{\mu_1}}\left(\sum_{k=0}^{d-1} V(X_k)\right) + 2\sum_{n=1}^{\infty} Cov_{\mathbb{P}_{\mu_1}}\left(\sum_{k=0}^{d-1} V(X_k), \sum_{k=0}^{d-1} V(X_{nd+k})\right).\end{aligned}$$

From the cyclic decomposition, we see that the last quantity above is exactly $\sigma^2(V)$. Thus $\Lambda''(tV)|_{t=0} = \sigma^2(V)$.

(b) By the LDP of L_n in [5] and the Laplace principle due to Varadhan, we have for all $t \in \mathbb{R}$,

$$\Lambda(t[V-\mu(V)]) = \sup\{\nu(tV) - t\mu(V) - J(\nu);\ \nu \in M_1(E)\} = \sup_{r\in\mathbb{R}}\{tr - J_V(r)\},$$

Hence the Legendre-Fenchel theorem gives us

$$J_V(r) = \sup_{t\in\mathbb{R}}\{tr - \Lambda(t[V-\mu(V)])\} = \sup_{t\in\mathbb{R}}\{t(r+\mu(V)] - \Lambda(tV)\},\ \forall r \in \mathbb{R}$$

for $\Lambda(t[V-\mu(V)]) = \Lambda(tV) - t\mu(V)$. When $r \ge 0$, since $\frac{d}{dt}\Lambda(tV)|_{t=0} = \mu(V)$, the supremum above can be taken only for $t \ge 0$. Then (3.28) is proved.

All other properties of $J_V(r) = \sup_{t\in\mathbb{R}}(tr - \Lambda(t[V-\mu(V)]))$ are easy consequences of the elementary convex analysis. □

Lemma 3.12 *(Bishop-Phelps)(cf.*[20] or [22]) Assume Λ is a convex real function on a Banach space Y. Assume $x_o \in Y'$ (the topological dual space) satisfies:

$$\exists c \in \mathbb{R} : \Lambda(y) \ge \langle x_0, y\rangle - c, \quad \forall y \in Y$$

then $\forall y \in Y, \forall \varepsilon > 0, \exists y' \in Y, x' \in \partial\Lambda(y')$, such that

$$\|x' - x_0\| \le \varepsilon, \|y'-y\| \le \frac{1}{\varepsilon}(\Lambda(y) - \langle x_0, y\rangle + \Lambda^*(x_0))$$

where $\Lambda^*(x) := \sup\{\langle x, y\rangle - \Lambda(y) \mid y \in Y\}),\ \forall x \in Y'$ is the Legendre transformation of $\Lambda(y)$.

3.4 Proof of Theorem 3.1

The desired LDP of f_n^* in $(L^1(\mathbb{R}^d), \sigma(L^1, L^\infty))$ is equivalent to the LDP of $f_n^*(x)dx$ on $M_1(\mathbb{R}^d)$ w.r.t. the τ-topology $\sigma(M_1(\mathbb{R}^d), b\mathcal{B})$. Since $\Lambda(V1_E)$ is Gateaux-differentiable on $b\mathcal{B}$ by Lemma 3.7(a), by the abstract Gärtner-Ellis theorem ([22], p290, Theorem 2.7), it is enough to show that for each $V \in b\mathcal{B}$,

$$\begin{aligned} &\liminf_{n\to\infty} \frac{1}{n} \log \inf_{x\in E} \mathbb{E}^x \exp\left(n \int_{\mathbb{R}^d} f_n^*(y)V(y)dy\right) \\ &= \limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in E} \mathbb{E}^x \exp\left(n \int_{\mathbb{R}^d} f_n^*(y)V(y)dy\right) = \Lambda(V1_E) \end{aligned} \tag{3.29}$$

and $\Lambda(V1_E)$ is monotonely continuous at 0, i.e., if (V_n) is a sequence in $b\mathcal{B}$ decreasing pointwise to 0 over $\mathbb{R}^d$, then $\Lambda(V_n1_E) \to 0$.

The last condition is satisfied by Lemma 3.7(b). It remains to verify (3.29). Put $V_n = (K_{h_n} * V)1_E$, then $\|V_n\| \le \|V\|$ and,

$$n \int_{\mathbb{R}^d} f_n^*(y)V(y)dy = \sum_{k=0}^{n-1} V_n(X_k).$$

Consequently letting ϕ_n be the right eigenfunction of P^{V_n} associated with $e^{\Lambda(V_n)}$, and $C := Me^{3N\|V\|}$, we have by Lemma 3.5(a) that for each $x \in E$,

$$\begin{aligned} \mathbb{E}^x \exp\left(n \int_{\mathbb{R}^d} f_n^*(y)V(y)dy\right) &\le C\mathbb{E}^x \frac{\phi_n(X_n)}{\phi_n(x)} \exp\left(\sum_{k=0}^{n-1} V_n(X_k)\right) \\ &= C\frac{(P^{V_n})^n\phi_n(x)}{\phi_n(x)} = Ce^{n\Lambda(V_n)} \end{aligned}$$

and similarly

$$\mathbb{E}^x \exp\left(n \int_{\mathbb{R}^d} f_n^*(y)V(y)dy\right) \ge \frac{1}{C}\mathbb{E}^x \frac{\phi_n(X_n)}{\phi_n(x)} \exp\left(\sum_{k=0}^{n-1} V_n(X_k)\right) = \frac{1}{C}e^{n\Lambda(V_n)}.$$

Noting that $V_n \to V1_E,\ dx - a.e.$, we have $\Lambda(V_n) \to \Lambda(V1_E)$ by Lemma 3.7(b). Thus the two estimations above yield the desired relation (3.29).

3.5 Proof of Theorem 3.2

Part 1. Large deviation upper bound. This is an easy consequence of Theorem 3.1. In fact, for any $g \in L^1(\mathbb{R}^d)$ and δ fixed, as $\{\tilde{g} \in L^1(\mathbb{R}^d);\ \|\tilde{g} - g\|_1 \le \delta\}$ is closed in the weak topology $\sigma(L^1, L^\infty)$, then by Theorem 3.1,

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in E} \mathbb{P}_x(\|f_n^* - g\|_{L^1(\mathbb{R}^d)} \le \delta) \le -\inf_{\tilde{g}:\|\tilde{g}-g\|_1\le\delta} J(\tilde{g}).$$

Letting $\delta \to 0$, we get the desired result by the lower semi-continuity of J (which follows from (3.18)).

Part 2. Large deviation lower bound. It is enough to prove that for any $g \in \mathcal{P}$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \inf_{x\in E} \mathbb{P}_x(\|f_n^* - g\|_1 < \delta) \ge -J(g), \quad \forall \delta > 0.$$

Its proof, more difficult, is divided into three steps.

Step 1. We claim that it is enough to show that for any $g \in \mathcal{P}$ and $\delta > 0$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x(\|f_n^* - g\|_1 < \delta) \ge -J(g),\ \ \mu - a.s.\ x \in A \tag{3.30}$$

for some $A \in \mathcal{B}(E)$ charged by μ. Indeed, if (3.30) is true, then by Egorov's lemma, there is some measurable $U \subset A$ with $\mu(U) > 0$ such that

$$\liminf_{n\to\infty} \frac{1}{n} \log \inf_{x\in U} \mathbb{P}_x(\|f_n^* - g\|_1 < \delta) \ge -J(g).$$

Let $\tau_U := \inf\{n \ge 1;\ X_n \in U\}$ be the first hitting time to U. By **(H)**, we have $(1/N)\sum_{k=1}^N P^k(x, \cdot) \ge (1/M)\mu(\cdot)$, then

$$\inf_{x\in E} \mathbb{P}_x(\tau_U \le N) \ge \inf_{x\in E} \mathbb{E}^x \frac{\sum_{k=1}^N 1_U(X_k)}{N} \ge \frac{\mu(U)}{M} > 0.$$

Since

$$f_n^* \circ \theta^k := \frac{1}{n} \sum_{i=k}^{k+n-1} \frac{1}{h_n^d} K\left(\frac{x - X_i}{h_n}\right)$$

we have $\|f_n^* - f_n^* \circ \theta_{\tau_U}\|_1 \le \dfrac{2N}{n}$ on $[\tau_U \le N]$. Thus by the strong Markov property, we have for $n \ge N$ such that $2N/n < \delta/2$,

$$\inf_{x\in E} \mathbb{P}_x(\|f_n^* - g\|_1 < \delta) \ge \inf_{x\in E} \mathbb{P}_x(\tau_U \le N) \cdot \inf_{y\in U} \mathbb{P}_y\left(\|f_n^* - g\|_1 < \frac{\delta}{2}\right)$$

where the desired uniform lower bound follows from (3.30).

Step 2. For "$gdx = \nu_V$"case. The idea of this step is to use change of measure. Given $V \in b\mathcal{B}$, let Q^V be the transition kernel defined in Lemma 3.5 and $\nu_V = \phi\psi\mu$. From Lemma 3.5, we know that Q^V is Doeblin recurrent.

Let $\mathbb{Q}^V_{\omega(0)}$ be the law of the Markov process with transition kernel Q^V and the initial point $\omega(0)$, which is $\nu_V - a.s.$ well defined on $\Omega = E^{\mathbb{N}}$, and $\mathbb{Q}^V := \int \mathbb{Q}^V_{\omega(0)} d\nu_V(\omega(0))$. Denoting by $\xi(\omega)$ the density of $\mathbb{Q}^V_{\omega(0)}$ w.r.t $\mathbb{P}_{\omega(0)}$ on $\sigma(X_1)$, we have for $\mu - a.s$ $\omega(0)$,

$$\left.\frac{d\mathbb{Q}^V_{\omega(0)}(d\omega_1, \cdots, d\omega_n)}{d\mathbb{P}_{\omega(0)}}\right|_{\mathcal{F}_n} = \exp\left(\sum\nolimits_{k=0}^{n-1} \log \xi(\theta^k \omega)\right)$$

and $\mathbb{E}^{\mathbb{Q}^V} \log \xi = J^{(2)}(\mathbb{Q}^V|_{\mathcal{F}_1}) = J(\nu_V)$ by Lemma 3.8. For any $\varepsilon > 0$, putting

$$W_n := \{\omega : \|f_n^*(\omega) - g\|_1 < \delta\}, \quad D_{n,\varepsilon} := \{\omega : \frac{1}{n}\sum_{n=0}^{n-1} \log \xi(\theta^k \omega) \le J(\nu_V) + \varepsilon\},$$

we have for $\mu - a.s$ $\omega(0)$,

$$\begin{aligned} \mathbb{P}_{\omega(0)}(W_n) &\ge \int_{W_n} \exp\left(-\sum_{k=0}^{n-1} \log \xi(\theta^k \omega)\right) d\mathbb{Q}^V_{\omega(0)} \\ &\ge \exp[-n(J(\nu_V) + \varepsilon)] \cdot \mathbb{Q}^V_{\omega(0)}\left(W_n \bigcap D_{n,\varepsilon}\right). \end{aligned} \tag{3.31}$$

So to get (3.30), it remains to show that $\mathbb{Q}^V_{\omega(0)}(D_{n,\varepsilon}) \to 1$ and $\mathbb{Q}^V_{\omega(0)}(W_n) \to 1$ for $\mu - a.s$ $\omega(0)$, as n goes to infinity (for any $\varepsilon > 0$).

By the ergodic theorem and the Fubini theorem, we have for $\nu_V \sim \mu - a.s$ $\omega(0)$,

$$\frac{1}{n}\sum_{k=0}^{n-1} \log \xi(\theta_k \omega) \to \mathbb{E}^{\mathbb{Q}^V} \log \xi = J(\nu_V), \quad \mathbb{Q}^V_{\omega(0)} - a.s.$$

where follows $\mathbb{Q}^V_{\omega(0)}(D_{n,\varepsilon}) \to 1$. For the second limit, applying the crucial Corollary 3.10 to $((X_n), \mathbb{Q}^V)$ (where the condition is satisfied because $((X_n), \mathbb{Q}^V)$ is Doeblin recurrent by Lemma 3.5), we have

$$\mathbb{Q}^V(W_n^c) \to 0 \text{ exponentially rapidly.}$$

Then by the Borel-Cantelli Lemma,

$$\mathbb{Q}^V(W_n^c, \text{infinitely often}) = 0.$$

By Fubini's theorem, $\mathbb{Q}^V_{\omega(0)}(W_n^c, \text{infinitely often}) = 0$, for $\nu_V \sim \mu$-a.s. $\omega(0)$.

Step 3. The general case.

By Step 1 and Step 2, it remains to show the :

Claim: $\forall \nu = gdx \in M_1(\mathbb{R}^d)$ *satisfies* $J(g) < +\infty$, *there exists a sequence of* $(\nu_n) := (\nu_{V_n})$, *such that* $\|\nu_{V_n} - \nu\|_{TV} \to 0$ *and* $\limsup_{n\to\infty} J(\nu_{V_n}) \le J(\nu)$.

Let us construct this sequence by means of Bishop-Phelps theorem (Lemma 3.12). For any $n \ge 1$, we choose $\tilde{V}_n \in b\mathcal{B}(E)$ such that $J(\nu) < \langle \nu, \tilde{V}_n\rangle - \Lambda(\tilde{V}_n) + \dfrac{1}{n}$ (by (3.18)). By Lemma 3.12, for each $\tilde{V}_n$ and $\varepsilon_n = \dfrac{1}{n(\|\tilde{V}_n\| + 1)}$, we can find $V_n \in b\mathcal{B}(E)$, $\nu_{V_n} \in \partial\Lambda(V_n)$ (which is a singleton $\{\nu_{V_n}\}$ by the proof of Lemma 3.8), such that

$$\|\nu_{V_n} - \nu\|_{TV} \le \varepsilon_n, \|\tilde{V}_n - V_n\| \le \frac{1}{\varepsilon_n}(\Lambda(\tilde{V}_n) - \langle \nu, \tilde{V}_n\rangle + J(\nu)).$$

So

$$\langle \nu_{V_n} - \nu, V_n\rangle \le \|\nu_{V_n} - \nu\|_{TV} \cdot \|V_n - \tilde{V}_n\| + \|\nu_{V_n} - \nu\|_{TV} \cdot \|\tilde{V}_n\| \le \frac{2}{n}.$$

As $\partial\Lambda(V_n) = \{\nu_{V_n}\}$, we have,

$$J(\nu_{V_n}) = \langle \nu_{V_n}, V_n\rangle - \Lambda(V_n) = \langle \nu_{V_n} - \nu, V_n\rangle + \langle \nu, V_n\rangle - \Lambda(V_n) \le \frac{2}{n} + J(\nu).$$

This proves the claim. The proof of the theorem is completed.

3.6 Proof of Theorem 3.3

3.6.1 Proof of part (a) in Theorem 3.3

Its proof is divided into two parts.

Part 1. Lower bound in (3.8). The lower bound is an easy consequence of Theorem 3.1. Actually, as $\{g \in L^1(\mathbb{R}^d);\ \|g - f\|_1 > \delta\}$ is open in the weak topology $\sigma(L^1, L^\infty)$, we have by Theorem 3.1,

$$\liminf_{n\to\infty} \frac{1}{n} \log \inf_{x\in E} \mathbb{P}_x(\|f_n^* - f\|_1 > \delta) \ge - \inf_{g:\ \|g-f\|_1 > \delta} J(g) = -I(\delta).$$

Part 2. Upper bound in (3.8). The proof of the upper bound is much more difficult, and it is divided into three steps, where the first two steps are similar to [6] and the third one is inspired by [16].

Step 1 (Approximation of K) The purpose of this step is to show that we can reduce to the case where $K = \frac{1}{|A|}1_A$, $A := \prod_{i=1}^d [x_i, x_i + a_i)$ is a rectangle (here $|A|$ denotes the Lebesgue measure of $A \in \mathcal{B}$).

Given $\varepsilon > 0$, we can find finite positive constants q, m, $b_1, \cdots, b_m$ and disjoint finite rectangles $A_1, \cdots, A_m$ in $\mathbb{R}^d$ of form $\prod_{i=1}^d [x_i, x_i + a_i)$ such that the function

$$K^{(\varepsilon)}(x) = \sum_{j=1}^m b_j I_{A_j}(x)$$

satisfies: $\int K^{(\varepsilon)}(x)dx = 1$, $K^{(\varepsilon)} \le q$ and $\int |K(x) - K^{(\varepsilon)}|dx < \varepsilon$. Define

$$f_n^{(\varepsilon),*} := K_{h_n}^{(\varepsilon)} * dL_n = \frac{1}{n} \sum_{k=0}^{n-1} \frac{1}{h_n^d} K^{(\varepsilon)} \left(\frac{x - X_k}{h_n} \right).$$

Then

$$\begin{aligned} \int |f_n^*(x) - f_n^{(\varepsilon),*}(x)|dx &\le \int h_n^{-d} \int |K^{(\varepsilon)}(\frac{x-y}{h_n}) - K(\frac{x-y}{h_n})| L_n(dy)dx \\ &= \int_{\mathbb{R}^d} |K^\varepsilon - K|(z)dz \le \varepsilon \end{aligned}$$

Thus by the approximation lemma in large deviations [4] (more precisely, by the same cycle of idea), it is enough to prove that $f_n^{(\varepsilon),*}$ satisfies the upper bound in (3.8).

Let $K^j = \frac{1}{|A_j|} 1_{A_j}$, then $K^{(\varepsilon)} = \sum_{j=1}^m \lambda_j K^j$ where $\sum_{j=1}^m \lambda_j = 1$ and $\lambda_j > 0$. Consequently,

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in E} \mathbb{P}_x \left(\|f_n^{(\varepsilon),*} - f\|_1 > \delta \right) \\ &\le \limsup_{n\to\infty} \frac{1}{n} \log \sum_{j=1}^m \sup_{x\in E} \mathbb{P}_x \left(\|K_{h_n}^j * dL_n - f\|_1 > \delta \right) \\ &= \max_{1\le j\le m} \limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in E} \mathbb{P}_x \left(\|K_{h_n}^j * dL_n - f\|_1 > \delta \right). \end{aligned}$$

Thus for the upper bound in (3.8), we may (and will) assume that $K = \frac{1}{|A|} 1_A$ where $A := \prod_{i=1}^d [x_i, x_i + a_i)$.

Step 2. (the method of partition) Fix such a rectangle $A := \prod_{i=1}^d [x_i, x_i + a_i)$ and $K = \frac{1}{|A|} 1_A$, and let $0 < \varepsilon < \delta/4$ be arbitrary. Since $K_{h_n} * f \to f$ in $L^1(\mathbb{R}^d)$, then it is enough to show that

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in E} \mathbb{P}_x \left(\|f_n^* - K_{h_n} * f\|_1 > \delta \right) \le -I(\delta-). \tag{3.32}$$

Note that

$$\begin{aligned}\int |f_n^*(x) - K_{h_n} * f(x)|dx &\leq \int \left| \frac{1}{|A|h_n^d} \int_{x+h_nA} L_n(dy) - \frac{1}{|A|h_n^d} \int_{x+h_nA} f(y)dy \right| dx \\ &\leq \frac{1}{|A|h_n^d} \int |L_n(x+h_nA) - \mu(x+h_nA)|dx.\end{aligned}$$

Consider the partition of $\mathbb{R}^d$ into sets B that are d-fold products of intervals of the form $[\frac{(i-1)h_n}{p}, \frac{ih_n}{p})$, where $i \in \mathbb{Z}$, and $p \in \mathbb{N}^*$ such that $\min_i a_i \geq \frac{2}{p}$. Call the partition Ψ.

Let $A^* = \Pi_{i=1}^d [x_i + \frac{1}{p}, x_i + a_i - \frac{1}{p})$. We have

$$C_x := (x + h_nA) \backslash \bigcup_{B \in \Psi, B \subseteq x+h_nA} B \subseteq x + h_n(A \backslash A^*).$$

Consequently,

$$\begin{aligned}&\int |f_n^*(x) - K_{h_n} * f(x)|dx \\ &\leq \frac{1}{|A|h_n^d} \int \sum_{B \in \Psi, B \subseteq x+hA} |L_n(B) - \mu(B)|dx + \frac{1}{|A|h_n^d} \int \{\mu(C_x) + L_n(C_x)\}dx.\end{aligned} \tag{3.33}$$

Using the fact that for any set $C \in \mathcal{B}$, $h > 0$ and any probability measure ν on $\mathbb{R}^d$,

$$\int \nu(x+hC)dx = |hC| = h^d|C|$$

(by Fubini), the last term in (3.33) is bounded from above by

$$\begin{aligned}\frac{1}{|A|h_n^d} 2h_n^d |A \backslash A^*| &= \frac{2}{|A|} \left(\prod_{i=1}^d a_i - \prod_{i=1}^d \left(a_i - \frac{2}{p} \right) \right) \\ &= 2 \left(1 - \prod_{i=1}^d \left(1 - \frac{2}{pa_i} \right) \right) \leq \varepsilon\end{aligned}$$

once if p verifies

$$\min_i a_i \geq \frac{2}{p}, \text{ and } 2 \left(1 - \prod_{i=1}^d \left(1 - \frac{2}{pa_i} \right) \right) \leq \varepsilon.$$

We fix such p which is independent of n.

For any finite constant $R>0$, letting $S_{OR} := \{x \in \mathbb{R}^d;\ |x| \le R\}$, we can bound the first term at the r.h.s. of (3.33) from above by

$$\sum_{B\in\Psi, B\cap S_{OR}\neq\phi} |L_n(B)-\mu(B)| \frac{1}{|A|h_n^d}\int_{B\subseteq x+h_nA} dx \\ + \frac{1}{|A|h_n^d}\int_{B\subseteq x+h_nA} dx\{L_n(S_{OR}^c)-\mu(S_{OR}^c)+2\mu(S_{OR}^c)\}.$$

Clearly, $h_n^{-d}\int_{B\subseteq x+h_nA} dx \le |A|$, and $\mu(S_{OR}^c) < \varepsilon/2$ for all $R \ge R_0$.

By Lemma 3.6, we have for all $t>0$,

$$\limsup_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{P}_x\{L_n(S_{OR}^c)-\mu(S_{OR}^c)>\varepsilon\} \\ \le -J_{S_{OR}^c}(\varepsilon) \le -\left(t[\varepsilon+\mu(S_{OR}^c)] - \Lambda(t1_{S_{OR}^c})\right).$$

Since $\lim_{R\to\infty}\Lambda(t1_{S_{OR}^c}) = 0$ by Lemma 3.7, then for any $L>0$, the l.h.s. above is bounded from above by $-L$ for all R large enough, say $R \ge R_1$. Fix $R \ge R_0 \vee R_1$ below. Summarizing those estimations we obtain

$$\limsup_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{P}_x\left(\int |f_n^*(x)-K_{h_n}*f(x)|dx > \delta\right) \\ \le (-L)\vee\limsup_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{P}_x\left(\sum_{B\in\Psi, B\cap S_{OR}\neq\emptyset}|L_n(B)-\mu(B)| > \delta-3\varepsilon\right). \tag{3.34}$$

Step 3. It remains to control the last term in (3.34). Set

$$\tilde{\Psi} = \{B; B\in\Psi, B\cap S_{OR}\neq\emptyset\}\bigcup\{C\},\ C := \left(\bigcup_{B\in\tilde{\Psi}} B\right)^c$$

and $\mathcal{B}(\tilde{\Psi}) = \sigma\{B; B\in\tilde{\Psi}\}$, the σ-field generated by $\tilde{\Psi}$. Regarding L_n and μ as probability measures on $\mathcal{B}(\tilde{\Psi})$, and denoting the total variation of $L_n-\mu$ on $\mathcal{B}(\tilde{\Psi})$ by $\|L_n-\mu\|_{\mathcal{B}(\tilde{\Psi})}$, we have

$$\sum_{B\in\Psi, B\cap S_{OR}\neq\emptyset}|L_n(B)-\mu(B)| \le \|L_n-\mu\|_{\mathcal{B}(\tilde{\Psi})} = \max_{V\in\{-1,1\}^{\tilde{\Psi}}}(L_n(V)-\mu(V))$$

where $\{-1,1\}^{\tilde{\Psi}}$ denotes the set of all $\mathcal{B}(\tilde{\Psi})$-measurable functions with values in $\{-1,1\}$ (which can be identified as the set of functions from $\tilde{\Psi}$ to $\{-1,1\}$). Therefore, for any $r > 0$ fixed,

$$\begin{aligned}\mathbb{P}_x\left(\sum_{B\in\Psi, B\cap S_O R\neq\emptyset}|L_n(B)-\mu(B)|>r\right) &\le \mathbb{P}_x\left(\max_{V\in\{-1,1\}^{\mathcal{B}(\tilde\Psi)}} L_n(V)-\mu(V)>r\right)\\ &\le \sum_{V\in\{-1,1\}^{\mathcal{B}(\tilde\Psi)}} \mathbb{P}_x\left(L_n(V)-\mu(V)>r\right).\end{aligned}$$

At first by Lemma 3.6, for each $V \in \{-1,1\}^{\tilde\Psi}$ and for all $0<\varepsilon<r$,

$$\sup_{x\in E}\mathbb{P}_x\left(L_n(V)-\mu(V)>r\right)\le M\exp\left(-nJ_{V1_E}(r-\varepsilon)\right),\ \forall n\ge\frac{4N}{\varepsilon}.$$

Secondly, the number of elements $\tilde\Psi$ is not greater than $\left(\dfrac{2Rp}{h_n}+2\right)^d+1=o(n)$ by (3.3), then $\{-1,1\}^{\tilde\Psi}$ has $2^{o(n)}$ elements for n large enough. Consequently letting $\mathbb{B}(1)$ be the unit ball in $L^\infty(\mu)$, we have

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{P}_x\left(\sum_{B\in\Psi, B\cap S_O R\neq\emptyset}|L_n(B)-\mu(B)|>r\right)\\ &\le\limsup_{n\to\infty}\frac{1}{n}\log 2^{o(n)}M\sup_{V\in\mathbb{B}(1)}\exp\left(-nJ_V(r-\varepsilon)\right)\\ &=-\inf_{V\in\mathbb{B}(1)}J_V(r-\varepsilon)\end{aligned}$$

where it follows by (3.34),

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{P}_x\left(\int|f_n^*(x)-K_{h_n}*f(x)|dx>\delta\right)\\ &\le(-L)\vee\left(-\inf_{V\in\mathbb{B}(1)}J_V(\delta-4\varepsilon)\right)\end{aligned}$$

As $L,\varepsilon>0$ are arbitrary and $\lim_{\varepsilon\to0+}\inf_{V\in\mathbb{B}(1)}J_V(\delta-4\varepsilon)=I(\delta-)$ by (3.13), we obtain the desired (3.33) and then complete the proof of the upper bound in (3.8).

3.6.2 Proof of Part (b) in Theorem 3.3

Let $J(\nu/P)$ be the Donsker-Varadhan entropy of ν w.r.t. the Markov kernel P given by (3.5). We have for any $1\le u\in b\mathcal{B}(E)$,

$$\int\log\frac{u}{P^Nu}d\nu=\sum_{k=0}^{N-1}\int\log\frac{P^ku}{PP^ku}d\nu\le NJ(\nu/P).$$

We get thus

$$NJ(\nu/P) \geq \sup_{1\leq u\in b\mathcal{B}(E)} \int \log \frac{u}{P^N u} d\nu = J(\nu/P^N), \ \forall \nu \in M_1(E), \forall N \geq 1. \quad (3.35)$$

By **(H)**, $P^l(x, \cdot) \leq M\mu(\cdot)$. Then

$$J(\nu/P) \geq \frac{J(\nu/P^l)}{l} \geq \frac{1}{l}\left(\sup_{1\leq u\in b\mathcal{B}(E)} \int \log \frac{u}{\mu(u)} d\nu - \log M\right) = \frac{h(\nu/\mu) - \log M}{l}$$

where $h(\nu/\mu) = \int \log \frac{d\nu}{d\mu} d\nu$ if $\nu \ll \mu$ and $+\infty$ otherwise, is the relative entropy of ν w.r.t μ (the last equality is the famous variational formula of relative entropy). Notice that in the i.i.d. case of common law μ, its transition is $P_0 f = \mu(f)$ and $h(\nu/\mu) = J(\nu/P_0)$. Hence

$$I^{iid}(\delta) = \inf\{h(\nu/\mu); \ \|\nu - \mu\|_{TV} > \delta\}$$

where the desired inequality (3.10) follows.

3.6.3 Proof of Part (c) in Theorem 3.3.

This follows from Rio's deviation inequality [19]. In fact using his inequality, we have (see [15] for details)

$$\mathbb{P}_\mu\left(\|f_n^* - f\|_1 - \mathbb{E}^\mu\|f_n^* - f\|_1 > \delta\right) \leq \exp\left(-\frac{n\delta^2}{2(1+2S_\phi)^2}\right), \ \forall n \geq 1, \delta > 0$$

where $S_\phi := \sum_{k=1}^{+\infty} \phi_k$ where ϕ_k is the ϕ-uniform mixing coefficient given in [19] or [15]. In the actual Markov context, we have

$$2\phi_k \leq \sup_{x,y\in E} \|P^k(x,\cdot) - P^k(y,\cdot)\|_{TV}$$

and then $2S_\phi \leq S$, the quantity in (3.11). S is finite for aperiodic Doeblin recurrent Markov chain. Moreover by Lemma 3.9, $\mathbb{E}^\mu\|f_n^* - f\|_1 \to 0$. Thus by the lower bound in (3.8), Rio's estimate and the right continuity of $I(\delta)$, we get

$$-I(\delta) \leq -\frac{\delta^2}{2(1+S)^2}$$

where the desired inequality (3.11) follows.

3.7 Proof of Theorem 3.4

Lemma 3.13 *Given $V \in b\mathcal{B}(E)$. If T_n is an asymptotically consistent estimator of $\langle V, f\rangle := \int_E V(x)f(x)dx$, i.e., for each $(P,\mu) \in \Theta$ (satisfying* **(H)** *and $d\mu(x) \ll dx$), $|\langle T_n, V\rangle - \langle f, V\rangle| \to 0$ in probability $\mathbb{P}_\mu$, then*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu\left(\langle T_n - f, V\rangle > \delta\right) \geq -\inf\{J(g);\ \langle g - f, V\rangle > \delta\}. \tag{3.36}$$

Proof. It is enough to prove that the l.h.s. of (3.36) is $\geq -J(g)$ for every $g \in \mathcal{P}(E)$ such that $\langle g - f, V\rangle > \delta$ and $J(g) < +\infty$. By the Step 3 of the proof of Theorem 3.2, it suffices to prove it for $gdx = \nu_{\tilde{V}}$ where $\tilde{V} \in b\mathcal{B}(E)$ is arbitrary. Its proof, completely parallel to the Step 2 in the proof of Theorem 3.2, is based on the fact that $(Q^{\tilde{V}}, \nu_{\tilde{V}}) \in \Theta$ again. It is so omitted. □

Lemma 3.14 *Under* **(H)**, *let $I(\cdot)$ be defined in (3.9). Then*

$$\lim_{r\to 0+} \frac{I(r)}{r^2} = \frac{1}{2\sup_{\|V\|\leq 1} \sigma^2(V)} = \frac{1}{8\sup_{A\in\mathcal{B}(E)} \sigma^2(1_A)}. \tag{3.37}$$

Proof. We shall only prove the first equality in (3.37) (the proof of the second is similar). By (3.13) and Lemma 3.11(b), for any $V \in b\mathcal{B}(E)$ with $\|V\| \leq 1$,

$$\limsup_{r\to 0} \frac{I(r)}{r^2} \leq \lim_{r\to 0} \frac{J_V(r+)}{r^2} = \frac{1}{2\sigma^2(V)}$$

where "$\leq$" in the first equality of (3.37) follows.

For the inverse inequality, let $L > 1$ be arbitrary but fixed. For any $\delta > 0$ small enough, we have by Lemma 3.11,

$$C(L\delta) := \sup_{t\in[0,L\delta]} \sup_{V\in\mathbb{B}(1)} \left|\frac{d^3}{dt^3}\Lambda(tV)\right| < +\infty.$$

Thus by the Taylor formula of order 3, we get for any $V \in \mathbb{B}(1)$ and $r \in (0, \delta]$,

$$\begin{aligned} J_V(r) &\geq \sup_{t\in[0,Lr]} \left(tr - \Lambda(t[V - \mu(V)])\right) \\ &\geq \sup_{t\in[0,Lr]} \left(tr - \frac{t^2\sigma^2(V)}{2}\right) - \frac{(Lr)^3}{6} \cdot C(L\delta) \\ &\geq r^2 \left(L \wedge \sigma^{-2}(V) - \frac{[L \wedge \sigma^{-2}(V)]^2\sigma^2(V)}{2}\right) - \frac{(Lr)^3}{6} \cdot C(L\delta) \end{aligned}$$

where the last inequality is obtained by taking $t = r[L \wedge \sigma^{-2}(V)]$. Thus by (3.13),

$$\begin{aligned}\liminf_{r\to 0+} \frac{I(r)}{r^2} &= \liminf_{r\to 0+} \inf_{V\in\mathbb{B}(1)} \frac{J_V(r)}{r^2} \\ &\geq \min\left\{ \inf_{V\in\mathbb{B}(1):\ \sigma^{-2}(V)\leq L} \frac{1}{2\sigma^2(V))};\ \inf_{V\in\mathbb{B}(1):\ \sigma^{-2}(V)>L} (L - L/2)\right\} \\ &\geq \min\left\{ \inf_{V\in\mathbb{B}(1)} \frac{1}{2\sigma^2(V)};\ \frac{L}{2}\right\}\end{aligned}$$

where the desired inverse inequality follows by letting $L \to +\infty$. □

Proof. [Proof of Theorem 3.4] (a) By Lemma 7.1, since $\mathcal{D}$ is dense in the unit ball of $L^\infty(\mathbb{R}^d)$ w.r.t. $\sigma(L^\infty, L^1)$,

$$\begin{aligned}&\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\|T_n - f\|_1 > r) = \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu\left(\sup_{V\in\mathcal{D}} \langle T_n - f, V\rangle > r\right) \\ &\geq \sup_{V\in\mathcal{D}} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\langle T_n - f, V\rangle > r) \\ &\geq -\inf_{V\in\mathcal{D}} \inf\{J(g) | \langle g - f, V\rangle > r\} = -\inf\{J(g) | \sup_{V\in\mathcal{D}} \langle g - f, V\rangle > r\} \\ &= -\inf_{g:\|g-f\|_1>r} J(g) = -I(r).\end{aligned}$$

Thus (3.15) follows from Lemma 3.14. The second claim follows easily from (3.15) by means of the extra condition on T_n and **(H)** (as in Step 1 of the proof of Theorem 3.2).

(b) It follows from Theorem 3.3 and Lemma 3.14. □

Bibliography

[1] Bosq, D., Merlevède, F. and Peligrad, M. (1999) Asymptotic normality for density kernel estimators in discrete and continuous time. *J. Multivariate Anal.*, **68**, 78-95.

[2] Carrasco, M. and Chen, X. (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, **18**, 17-39.

[3] Csörgö, M. and Horváth, L. (1988) Central limit theorems for L_p-norms of density estimators. *Z. Wahrsch. Verw. Gebiete.*, **80**, 269-291.

[4] Dembo, A. and Zeitouni, O. (1998) *Large Deviations Techniques and Applications.* 2nd Edn, Springer, New York.

[5] Deuschel, J.D. and Stroock, D.W. (1989) *Large Deviations.* Pure Appl. Math., vol 137, Academic Press, San Diego.

[6] Devroye, L. (1983) The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates. *Ann. Statist.*, **11**, 896-904.

[7] Donsker, M.D. and Varadhan, S.R.S. (1975; 1983) Asymptotic evaluation of certain Markov process expectations for large time, I-IV. *Commun. Pure Appl. Math.*, (1975) **28**, 1-47, 279-301; **29** 389-461; (1983) **36** , 183-212.

[8] Gao, F.Q., (2003) Moderate deviations and large deviations for kernel density estimators. *J. Theoret. Probab.*, **16**, 401-418.

[9] Gao, F.Q., Moderate deviations and the law of iterated logarithm for kernel density estimators in L_1. Preprint 2003.

[10] Giné, E., Mason, D.M. and Zaitsev, A.Yu. (2003) The L_1-norm density estimator process. *Ann. Probab.*, **31**, 719-768.

[11] Horváth, L. (1991) On L_p-norms of multivariate density estimators. *Ann. Statist.*, **19**, 1933-1949.

[12] Kato, T. (1992) *Perturbation Theory for Linear Operators.* 2nd Edn, Spring-Berlin, New York/Berlin.

[13] Kontoyiannis, I. and Meyn, S.P. (2003) Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.*, **13**, 304-362.

[14] Lei, L.Z., Wu, L.M. and Xie, B. (2003) Large deviations and deviation inequality for kernel density estimator in $L_1(R^d)$-distance. In *Development of Modern Statistics and Related Topics*, Ser. Biostat. 1., 89-97, *World Sci. Publishing, River Edge, NJ.*

[15] Lei, L.Z. and Wu, L.M. (2004) The exponential convergence of kernel density estimator in L_1 for ϕ-mixing processes. *Ann.I.S.U.P.*, **48**, 59-68.

[16] Louani, D. (2000) Large deviations for the L_1-distance in kernel density estimation. *J. Statist. Plann. Inference.*, 177-182.

[17] Meyn, S.P. and Tweedie, R.L. (1993) *Markov chains and stochastic stability.* Springer, Berlin.

[18] Peligrad, M. (1992) Properties of uniform consistency of the kernel estimators of density and of regression functions under dependence assumption. *Stochastics Stochastics Rep.*, **40**, 147-168.

[19] Rio, E. (2000) Inégalités de Hoeffding pour les fonctions lipchitziennes de suites dépendantes. *CRAS, Paris, t.330, Série I*, 905-908.

[20] Ruelle, D. (1978) *Thermodynamic Formalism*, Addison-Wesley, 1978.

[21] Wu, L.M. (1995) Moderate deviations of dependent random variables related to CLT. *Ann. Probab.*, **23**, 420-445.

[22] Wu, L.M. (1997) An introduction to large deviations. In: Yan, J.A., Peng, S.G., Fang, S.Z. and Wu, L.M. (Eds), *Several Topics in Stochastic Analysis*, 225-336, Academic Press of China, Beijing, (In Chinese).

[23] Wu, L.M. (2000) Uniformly integrable operator and large deviations for Markov processes. *J. Funct. Anal.*, **172**, 301-376.

Chapter 4

Large deviations of kernel density estimator in $L^1(\mathbb{R}^d)$ for reversible Markov processes

(To be published in: *Bernoulli*)

4.1 Introduction

Let $\{X_n; n \geq 0\}$ be a reversible $\mathbb{R}^d$-valued Markov chain, defined on the probability space $(\Omega, (\mathcal{F}_n^0)_{(n\in\mathbb{N})}, \mathcal{F}, (\mathbb{P}_x)_{x\in\mathbb{R}^d})$, with (unknown) Markov transition kernel $P(x, dy)$. Assume that

(A1) P is irreducible (Meyn and Tweedie 1993) and symmetric with respect to the unique invariant probability measure μ, which is absolutely continuous, i.e., $d\mu(x) = f(x)dx$, where the density f is unknown.

Given the observed sample $\{X_0, \dots, X_n\}$, consider the empirical measure of the ladder type, i.e.,

$$L_n = \frac{1}{n}\left(\sum_{i=1}^{n-1}\delta_{X_i} + \frac{1}{2}(\delta_{X_0} + \delta_{X_n})\right) = \frac{1}{n}\sum_{i=0}^{n-1}\frac{1}{2}\left(\delta_{X_i} + \delta_{X_{i+1}}\right),$$

Let $K : \mathbb{R}^d \to \mathbb{R}$ be a measurable function such that

$$K \geq 0, \int_{\mathbb{R}^d} K(x)dx = 1, \tag{4.1}$$

and set $K_h(x) = \frac{1}{h^d}K(\frac{x}{h})$ for any $h > 0$. The kernel density estimator of the unknown function f is defined as follows: for all $x \in \mathbb{R}^d$,

$$f_n^*(x) := K_{h_n} * dL_n(x) = \frac{1}{n}\sum_{i=0}^{n-1}\frac{1}{2h_n^d}\left(K\left(\frac{x-X_i}{h_n}\right)+K\left(\frac{x-X_{i+1}}{h_n}\right)\right), \tag{4.2}$$

where $h = h_n, \{h_n, n \geq 0\}$ is a sequence of positive numbers (bandwidth) satisfying

$$h_n \to 0, \quad nh_n^d \to +\infty \quad as \quad n \to \infty. \tag{4.3}$$

A natural distance of f_n^* from the unknown f is the L^1-distance,

$$D_n^* = \int_{\mathbb{R}^d} |f_n^*(x) - f(x)|dx. \tag{4.4}$$

The large deviation behavior of f_n^* in $(L^1(\mathbb{R}^d), \|\cdot\|_1 := \|\cdot\|_{L^1(\mathbb{R}^d)})$ is the subject of our study. In the i.i.d. case, due to Devroye (1983), all types of $L^1(\mathbb{R}^d)$-consistency of f_n^* are equivalent to condition (4.3) on the bandwidth. The asymptotic normality of D_n^* was investigated by Csörgö and Horváth (1988). Louani (2000) established the large deviation principle (LDP in short) for D_n^*, and recently Lei et al. (2003) proved the weak LDP of f_n^* in $L^1(\mathbb{R}^d)$, and showed that the corresponding LDP is false. More recently Gao (2003) obtained the moderate deviation principle of f_n^* in $L^1(\mathbb{R}^d)$ and the law of the iterated logarithm for D_n^*. Giné et al. (2003) established a functional central limit theorem and a Glivenko-Cantelli theorem for the density estimator process in L^1-norm.

A very natural question is how to extend those results from the i.i.d. case to the dependent case. Consistency of f_n^* has been studied by Peligrad (1992), Bosq et al. (1999), see also the references therein. But little is known about large deviations. Large deviation probabilities for f_n^* in L^1 and for D_n^* were obtained by Wu and Lei (2005) for uniformly ergodic Markov processes. Here uniform ergodicity means that there exist $1 \leq N \in \mathbb{N}^*$ and $C \geq 1$ such that

$$\frac{1}{C}\mu(\cdot) \leq \frac{1}{N}\sum_{k=1}^{N} P^k(x,\cdot) \leq C\mu(\cdot), \ \forall x \in E,$$

where E is a measurable subset of $\mathbb{R}^d$. The assumption is not satisfied by many discrete models with non-compact state space. For example, all real-valued stationary and ergodic Gaussian Markov processes are reversible but not uniformly ergodic. The purpose of this work is to establish the LDP for f_n^* in $L^1(\mathbb{R}^d)$ and for D_n^* in the framework of **(A1)** and **(A2)** below, instead of the strong 'uniform ergodicity' assumption.

(A2) For some $N \geq 1$, P^N is uniformly integrable in $L^2(\mu)$, i.e., $\{(P^N f)^2;\ \|f\|_{L^2(\mu)} \leq 1\}$ is uniformly integrable.

Wu (2000) proved that **(A2)** is a sufficient condition to obtain the LDP of L_n in the space $M_1(\mathbb{R}^d)$ of probability measures on $\mathbb{R}^d$ with respect to the τ-topology (this condition is even necessary in the reversible case, see Wu (2002)). The rate function is given by

$$J_\mu(\nu) := \begin{cases} \sup\left\{\int \log \frac{u}{Pu} d\nu;\ 1 \leq u \in b\mathcal{B}\right\}, & \forall \nu \in M_1(\mathbb{R}^d), \nu \ll \mu; \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.5}$$

where $b\mathcal{B}$ is the space of bounded and Borel-measurable functions on $\mathbb{R}^d$. We refer the reader to Wu (2000) for related references on the subject.

This paper is organized as follows: the main results are stated in the next section. In Section 3 we present several crucial lemmas which may have some independent interests, and we prove the main results in the rest of the paper.

4.2 Main results

Throughout this paper, we adopt the following notation:

$$L^p := L^p(\mathbb{R}^d) := L^p(\mathbb{R}^d, dx),\ \|f\|_p = \|f\|_{L^p(\mathbb{R}^d, dx)},\ L^p(\mu) := L^p(\mathbb{R}^d, \mu).$$

We denote for any $L \geq 1$,

$$\mathcal{A}_{\mu,2}(L) := \left\{\nu \in M_1(\mathbb{R}^d);\ \nu \ll \mu, \|\frac{d\nu}{d\mu}\|_{L^2(\mu)} \leq L\right\},\ \mathcal{A}_{\mu,2} := \bigcup_{L\geq 1} \mathcal{A}_{\mu,2}(L).$$

Throughout this paper we assume **(A1)** and **(A2)**.

When the bandwidth $h_n \to 0$, $f_n^* dx$ is 'close' to L_n in the τ-topology, one may hope that $f_n^* dx$ satisfies the same LDP as L_n. This intuition is in fact right:

Theorem 4.1 *Assume $h_n \to 0$ (without (4.3)). Then $\mathbb{P}_\nu(f_n^* \in \cdot)$ satisfies, uniformly over initial measures $\nu \in \mathcal{A}_{\mu,2}(L)$ for each $L \geq 1$, the LDP in L^1 w.r.t. the weak topology $\sigma(L^1, L^\infty)$ with the rate function*

$$J(g) := \begin{cases} J(gdx), & \text{if } gdx \in M_1(\mathbb{R}^d) \text{ and } gdx \ll fdx; \\ +\infty, & \text{otherwise,} \end{cases} \tag{4.6}$$

where $J(\cdot)$ is the Donsker-Varadhan entropy given in (4.5). More precisely, J is inf-compact on $(L^1, \sigma(L^1, L^\infty))$, and for any measurable subset A of L^1, for every $L \geq 1$,

$$
\begin{aligned}
-\inf_{g\in A^{o\sigma}} J(g) &\leq \liminf_{n\to\infty} \frac{1}{n}\log \inf_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(f_n^* \in A)\\
&\leq \limsup_{n\to\infty} \frac{1}{n}\log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(f_n^* \in A) \leq -\inf_{g\in\bar{A}^\sigma} J(g)
\end{aligned}
$$

where $A^{o\sigma}, \bar{A}^\sigma$ denote respectively the interior and the closure of A w.r.t. the weak topology $\sigma(L^1, L^\infty)$.

The LDP w.r.t. the weak topology on L^1 as above is too weak in the sense that it does not entail the consistency, i.e., $D_n^* \to 0$ in probability. For statistical issues, the main objects to be studied are

(i) $\mathbb{P}_\nu(\|f_n^* - g\|_1 < \delta)$ where $gdx \in M_1(\mathbb{R}^d)$ is fixed, which is important in the hypothesis testing: $H_0:\ d\mu(x) = f(x)dx$ against $H_1:\ d\mu(x) = g(x)dx$; or

(ii) $\mathbb{P}_\nu(D_n^* > \delta)$, whose statistical importance is obvious.

Unfortunately, Theorem 4.1 can not be applied to them, since $\{\tilde{g} \in L^1;\ \|\tilde{g}-g\|_1 < \delta\}$ is not open in $\sigma(L^1, L^\infty)$ and $\{\tilde{g} \in L^1;\ \|\tilde{g} - f\|_1 \geq \delta\}$ is not closed in $\sigma(L^1, L^\infty)$. Therefore, in order to deal with the objects (i) and (ii), we turn to Theorem 4.2 and Theorem 4.3 below.

Theorem 4.2 *Assume (4.3). Then for any $L \geq 1$ and for each $\delta > 0$,*

$$
\begin{aligned}
-I(\delta) &\leq \liminf_{n\to\infty} \frac{1}{n}\log \inf_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - f\|_1 > \delta)\\
&\leq \limsup_{n\to\infty} \frac{1}{n}\log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - f\|_1 \geq \delta) \leq -I(\delta-)
\end{aligned}
\tag{4.7}
$$

where

$$
I(\delta) = \inf\{J(g) | g \in L^1, \|g - f\|_1 > \delta\} > 0 \tag{4.8}
$$

and $I(\delta-)$ is the left-limit of I at δ.

Theorem 4.3 *Assume (4.3). Then $\mathbb{P}_\nu(f_n^* \in \cdot)$ satisfies weak*-LDP with rate function J on $(L^1, \|\cdot\|_1)$ uniformly over initial measures $\nu \in \mathcal{A}_{\mu,2}(L)$ for any $L \geq 1$, i.e., for any $L \geq 1$ and $g \in L^1$,*

$$
\begin{aligned}
&\lim_{\delta\to 0}\liminf_{n\to\infty} \frac{1}{n}\log \inf_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - g\|_1 < \delta)\\
&= \lim_{\delta\to 0}\limsup_{n\to\infty} \frac{1}{n}\log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - g\|_1 < \delta) = -J(g).
\end{aligned}
\tag{4.9}
$$

With the above results, we have established the deviation estimates of the estimator f_n^*, which are useful in statistics. Now, we claim that f_n^* is asymptotically optimal in the Bahadur sense. Let Θ be the set of unknown data (P,μ) satisfying **(A1)** and **(A2)**. Given a subset $\mathcal{D}$ of the unit ball in $b\mathcal{B}$, we say that an estimator $T_n(x) := T_n(x\ ; X_0, \cdots, X_n) \in L^1(\mathbb{R}^d, dx)$ is an **asymptotically** $\sigma(L^1,\mathcal{D})$**-consistent estimator** of the density f, if for all $V \in \mathcal{D}$, $\int_{\mathbb{R}^d} T_n(x)V(x)dx \to \int_{\mathbb{R}^d} f(x)V(x)dx$ in probability measure $\mathbb{P}_\mu$.

Theorem 4.4 *Given $(P,\mu) \in \Theta$, let $((X_n),(\mathbb{P}_x)_{x\in\mathbb{R}^d})$ be the associated Markov process.*

(a) **(Bahadur type lower bound)** *Assume that $\mathcal{D}$ is dense in the unit ball of L^∞ w.r.t. the weak* topology $\sigma(L^\infty, L^1)$. Then for any $\sigma(L^1,\mathcal{D})$-asymptotically consistent estimator T_n of the unknown density f,*

$$\begin{aligned}\liminf_{r\to 0+} \frac{1}{r^2} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\|T_n - f\|_1 > r)\\ \geq -\frac{1}{2\sup_{\|V\|\le 1}\sigma^2(V)} = -\frac{1}{8\sup_{A\in\mathcal{B}}\sigma^2(1_A)},\end{aligned} \tag{4.10}$$

where

$$\sigma^2(V) := 2\sum_{k=0}^{\infty}\langle V, P^k(V-\mu(V))\rangle_\mu - Var_\mu(V).$$

If moreover $\|T_n - T_n\circ\theta^N\|_1 \le \delta_n \to 0$, then (4.10) still holds with $\mathbb{P}_\mu$ substituted by $\mathbb{P}_\nu$ for any initial measure $\nu \in M_1(E)$, where θ is the shift on Ω.

(b) **(Asymptotic efficiency of f_n^* in the Bahadur sense)** *If h_n verifies (4.3), then*

$$\begin{aligned}&\liminf_{r\to 0+} \frac{1}{r^2} \liminf_{n\to\infty} \frac{1}{n} \log \inf_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - f\|_1 > r)\\ &= \limsup_{r\to 0+} \frac{1}{r^2} \limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - f\|_1 > r)\\ &= -\frac{1}{2\sup_{\|V\|\le 1}\sigma^2(V)} = -\frac{1}{8\sup_{A\in\mathcal{B}}\sigma^2(1_A)}.\end{aligned} \tag{4.11}$$

Thus f_n^* is an asymptotically efficient estimator of f in the Bahadur sense. And $1/\sigma^2(V)$ can be interpreted as the Fisher information in the direction V of our statistical model Θ.

4.3 Preliminary lemmas

For every $V \in b\mathcal{B}$, put $P^V(x,dy) := \exp\left(\frac{V(x)+V(y)}{2}\right) P(x,dy)$. We have the Feynman-Kac formula,

$$(P^V)^n f(x) = \mathbb{E}^{\mathbb{P}_x} f(X_n) \exp\left(\sum_{k=0}^{n-1} \frac{V(X_k)+V(X_{k+1})}{2}\right),$$

where $\mathbb{E}^{\mathbb{P}_x}$ is the expectation with respect to $\mathbb{P}_x$. Introduce the following Cramèr functional

$$\Lambda^{(2)}(V) := \lim_{n\to\infty} \frac{1}{n} \log \|(P^V)^n\|_{L^2(\mu)\to L^2(\mu)}, \tag{4.12}$$

then $e^{\Lambda^{(2)}(V)}$ is the spectral radius of P^V on $L^2(\mu)$. For the sake of convenience, we will write $\Lambda(V)$ for $\Lambda^{(2)}(V)$. It is well known that (see Wu 2000)

$$J_\mu(\nu) = \sup\{\nu(V) - \Lambda(V) | V \in b\mathcal{B}\}, \ \forall \nu \in M_1(\mathbb{R}^d). \tag{4.13}$$

On the other hand, by the continuity of Λ on $b\mathcal{B}$ w.r.t. the Mackey topology proved in Wu (2000, Theorem 5.1 and Theorem B.5) and by the Fenchel-Legendre theorem, we have for all $t \in \mathbb{R}$,

$$\Lambda(t[V-\mu(V)]) = \sup\{\nu(tV) - t\mu(V) - J_\mu(\nu);\ \nu \in M_1(E)\} = \sup_{r\in\mathbb{R}}\{tr - J_V(r)\}, \tag{4.14}$$

where $J_V(r)$ is given by

$$J_V(r) := \inf\{J_\mu(\nu);\ \nu \in M_1(E),\ \nu(V) = \mu(V) + r\}. \tag{4.15}$$

J_V is convex. By the LDP of $\mathbb{P}_\nu(L_n \in \cdot)$ in Wu (2000, Theorem 5.1) ($\nu \in \mathcal{A}_{\mu,2}$) and the contraction principle, $J_V : \mathbb{R} \to [0, +\infty]$ is inf-compact on $\mathbb{R}$ and $\mathbb{P}_\nu(L_n(V) - \mu(V) \in \cdot)$ verifies the LDP with the rate function J_V. Furthermore, by the Fenchel-Legendre theorem and (4.14), we have

$$J_V(r) = \sup_{t\in\mathbb{R}}\{tr - \Lambda(t[V-\mu(V)])\} = \sup_{t\in\mathbb{R}}\{t(r+\mu(V)] - \Lambda(tV)\},\ \forall r \in \mathbb{R}$$

for $\Lambda(t[V-\mu(V)]) = \Lambda(tV) - t\mu(V)$. When $r \geq 0$, the supremum above can be taken only for $t \geq 0$. Then we get

$$J_V(r) = \begin{cases} \sup_{t\in\mathbb{R}} \left(t[(r+\mu(V)] - \Lambda(tV)\right), & \forall r \in \mathbb{R}, \\ \sup_{t\geq 0} \left(t[(r+\mu(V)] - \Lambda(tV)\right), & \forall r \geq 0. \end{cases} \tag{4.16}$$

Part (b) in the lemma below is crucial and gives us a robust estimate which extends the well known inequality of Cramèr in the i.i.d. case.

Lemma 4.5 *For the positive operator* $P^V(x, dy) := \exp\left(\frac{V(x)+V(y)}{2}\right) P(x, dy)$,

(a) P^V *is also symmetric in* $L^2(\mu)$ *and* $\|P^V\|_{L^2(\mu)} = e^{\Lambda(V)}$, *and there exists* $\phi \in L^2(\mu)$ μ*-a.s. strictly positive such that* $\int_E \phi^2 d\mu = 1$ *and*

$$P^V\phi = e^{\Lambda(V)}\phi \quad over \ \ \mathbb{R}^d, \ \ \mu - a.s.$$

Moreover the eigenspace $Ker\left(e^{\Lambda(V)} - P^V\right)$ *of* P^V *associated with the eigenvalue* $e^{\Lambda(V)}$ *in* $L^2(\mu)$ *is spanned by* ϕ.

(b) *(A deviation inequality of Cramèr type) For any initial measure* $\nu \in \mathcal{A}_{\mu,2}$, $r > 0$,

$$\mathbb{P}_\nu\left(L_n(V) > \mu(V) + r\right) \le e^{-nJ_V(r)} \cdot \left\|\frac{d\nu}{d\mu}\right\|_{L^2(\mu)} \tag{4.17}$$

where $J_V(r) = \inf\{J_\mu(\nu); \nu(V) = \mu(V) + r\}$.

(c) *Define a Markov kernel* Q^V *as :*

$$Q^V(x, dy) = \frac{\phi(y)}{e^{\Lambda(V)}\phi(x)} e^{\frac{V(x)+V(y)}{2}} P(x, dy)$$

then $\nu_V := \phi^2\mu$ *is the unique invariant probability measure for* Q^V *and* Q^V *is symmetric on* $L^2(\nu_V)$.

The Cramèr type inequality (4.17) was established by Wu (2000b) in the continuous time case.

Proof. (a) Under **(A1)** and **(A2)**, P^V is again symmetric, uniformly integrable and irreducible on $L^2(\mu)$. Thus part (a) follows by Wu (2000, Theorem 3.1 and Corollary 3.3).

(b) By the symmetry of P^V on $L^2(\mu)$, we have $\|(P^V)^n\|_{L^2(\mu)} := \|(P^V)^n\|_{L^2(\mu)\to L^2(\mu)} = e^{n\Lambda(V)}$ for each $V \in b\mathcal{B}$. Thus for any initial measure $\nu \in \mathcal{A}_{\mu,2}$, $0 \le f \in L^2(\mu)$ and any $t \in \mathbb{R}$,

$$\mathbb{E}^\nu\left(f(X_n)e^{ntL_n(V)}\right) \le \left\|\frac{d\nu}{d\mu}\right\|_{L^2(\mu)} \cdot \|f\|_{L^2(\mu)} \cdot \|(P^{tV})^n\|_{L^2(\mu)} = \left\|\frac{d\nu}{d\mu}\right\|_{L^2(\mu)} \cdot \|f\|_{L^2(\mu)} \cdot e^{n\Lambda(tV)}. \tag{4.18}$$

By Chebychev's inequality,

$$\begin{aligned}\mathbb{E}^\nu\left(1_{[L_n(V)>\mu(V)+r]}f(X_n)\right) &\le \left\|\frac{d\nu}{d\mu}\right\|_{L^2(\mu)} \cdot \|f\|_{L^2(\mu)} \cdot \inf_{t\ge 0} e^{-nt(\mu(V)+r)+n\Lambda(V)} \\ &= \left\|\frac{d\nu}{d\mu}\right\|_{L^2(\mu)} \cdot \|f\|_{L^2(\mu)} \cdot e^{-nJ_V(r)}\end{aligned}$$

where the second equality follows from (4.16). So (4.17) holds.

(c) It is easy to verify that Q^V is a Markov kernel, that $\nu_V := \phi^2\mu$ is an invariant measure of Q^V, and that it is symmetric on $L^2(\nu_V)$. As Q^V is irreducible as well as P, $\phi^2\mu$ is the unique invariant measure of Q^V. $\square$

The following result is technically crucial for all the results in this paper.

Lemma 4.6 *(a) $\Lambda(V)$ is Gâteaux-differentiable on $b\mathcal{B}$.*

(b) If $V_n \to V$ in measure μ and $\sup_n \|V_n\| \le C$, then $\Lambda(V_n) \to \Lambda(V)$.

Proof. (a) Under **(A2)**, $(P^V)^N$ is uniformly integrable on $L^2(\mu)$, and P^V is irreducible. Thus by Wu (2000, Theorem 3.11), the largest eigenvalue $e^{\Lambda(V)}$ of P^V is isolated in the spectrum $\sigma(P^V)$ of P^V on $L^2(\mu)$, with simple algebraic multiplicity. Consequently, by the theory of perturbation of linear operators (Kato 1992, Chap.VII, Theorem 1.8), $e^{\Lambda(V)}$ is real-analytic on $b\mathcal{B}$, i.e., $\Lambda(V + t\tilde{V})$ is analytic on $t \in \mathbb{R}$ for any $V, \tilde{V} \in b\mathcal{B}$ fixed.

(b) First of all, $\liminf_{n\to\infty} \Lambda(V_n) \ge \Lambda(V)$ by (4.14). The converse inequality which is equivalent to $\limsup_{n\to\infty} e^{\Lambda(V_n)} \le e^{\Lambda(V)}$ follows by applying Wu (2000, Prop. 3.8) to $\pi_n := (P^{V_n})^N$. $\square$

Lemma 4.7 *(Gibbs type principle) Given a function $V \in b\mathcal{B}$, a probability measure $\nu \ll \mu$ on $\mathbb{R}^d$ satisfies*

$$J_\mu(\nu) = \langle \nu, V\rangle - \Lambda(V)$$

iff $\nu = \nu_V := \phi^2\mu$, where ϕ is the right eigenfunction of P^V associated with $e^{\Lambda(V)}$ given in Lemma 4.5(a) verifying $\mu(\phi^2) = 1$.

Proof. The proof is identical to that of Lei and Wu (2005, Lemma 3.4). $\square$

Lemma 4.8 *Under* **(A1)** *and* **(A2)***, for each $\nu = gdx \in M_1(\mathbb{R}^d)$ satisfying $J_\mu(\nu) < +\infty$, there exists a sequence of $(\nu_{V_n} = \phi_n^2 d\mu)$ given in Lemma 4.5, such that*

$$\|\nu_{V_n} - \nu\|_{TV} \to 0 \quad \textit{and} \quad \limsup_{n\to\infty} J(\nu_{V_n}) \le J_\mu(\nu).$$

Here $\|\cdot\|_{TV}$ means the total variation of a signed measure.

Proof. The proof is omitted; for details, we refer the reader to Lei and Wu (2005, Part 2, Proof of Theorem 2.2). $\square$

Lemma 4.9 *Under* **(A1)** *and* **(A2)**, *we have*

(a) For any $k \geq 1$, there exists some $\delta > 0$ such that

$$\sup_{|t|\leq\delta}\sup_{\|V\|\leq 1}\left|\frac{d^k}{dt^k}\Lambda(tV)\right| < +\infty$$

and for any $V \in b\mathcal{B}$, $\frac{d^2}{dt^2}\Lambda(tV)|_{t=0} = \sigma^2(V)$ which is given in Theorem 4.4.

(b) Let J_V be defined as in (4.15). Then J_V is strictly convex on $[J_V < +\infty]^0 = (a, b)$, where $a = \lim_{t\to-\infty}\frac{d}{dt}\Lambda(tV) - \mu(V)$ and $b = \lim_{t\to+\infty}\frac{d}{dt}\Lambda(tV) - \mu(V)$ (in particular, J_V is strictly increasing and continuous in $[0, b)$); moreover

$$\lim_{r\to 0+}\frac{J_V(r)}{r^2} = \frac{1}{2\sigma^2(V)} \in (0, +\infty].$$

Proof. (a) Under **(A1)** and **(A2)**, by Wu (2000, Theorem 3.11), 1 is an isolated point in the spectrum $\sigma(P)$ in $L^2(\mu)$ (i.e., the existence of spectral gap). We prove the lemma only in the case where $-1 \notin \sigma(P)$ (it corresponds to the aperiodicity of the irreducible chain). Otherwise, one may consider the periodic decomposition as in Lei and Wu (2005).

As in Wu (1995), we apply the analytical perturbation theory in Kato (1992). For each $z \in \mathbb{C}$, consider P^{zV} as an operator acting on the complexified space $L^2(E, \mu; \mathbb{C})$, which is analytical in z in the sense of Kato (1992). Then for any $\eta \in (0, 1/2)$ sufficiently small, there exist $\delta > 0$ and $C > 0$ such that for all $V \in b\mathcal{B}$ with $\|V\| \leq 1$,

1) the eigenvalue $\lambda_{max}(P^{zV})$ of P^{zV} with the largest modulus is isolated in $\sigma(P^{zV})$ and $|\lambda_{max}(P^{zV}) - 1| \leq \eta$ for $|z| \leq 2\delta$;

2) for all $|z| \leq 2\delta$, the eigenprojection $E(z, V)$ of P^{zV} associated with $\lambda_{max}(P^{zV})$ is unidimensional and

$$\|E(z, V)1 - 1\|_{L^2(\mu)} < 1/2,\ \|(P^{zV})^n(I - E(z, V))\|_{L^2(\mu)} \leq C(1 - 2\eta)^n,\ \forall n;$$

3) $z \to \lambda_{max}(P^{zV})$ and $z \to E(z, V)f$ are analytic in z for $|z| \leq 2\delta$ (for each $f \in L^2(\mu)$);

where properties 1) and 2) follow from Kato (1992, Chap.IV, Theorem 3.16), and the property 3) follows from Kato (1992, Chap.VII, Theorem 1.8).

Then $\Lambda(zV) := \log\lambda_{max}(P^{zV})$ is analytic for $|z| \leq 2\delta$ and coincides with $\Lambda(tV)$ when $z = t \in [-2\delta, 2\delta] \subset \mathbb{R}$.

Let $\Lambda_n(zV) := \frac{1}{n}\log\langle 1, (P^{zV})^n 1\rangle_\mu$. By properties 1) and 2) above, we have

$$\langle 1, (P^{zV})^n 1\rangle_\mu = e^{n\Lambda(zV)}\langle 1, E(z,V)1\rangle_\mu + O((1-2\eta)^n)$$

where it follows that $\Lambda_n(zV) \to \Lambda(zV)$ uniformly over $z:\ |z| \le 2\delta$ and $V : \|V\| \le 1$. Thus by Cauchy's theorem and property 3) above,

$$\sup_{\|V\|\le 1}\sup_{|z|\le\delta}\left|\frac{d^k}{dz^k}\Lambda(zV)\right| < +\infty,\ \ \sup_{\|V\|\le 1}\sup_{|z|\le\delta}\left|\frac{d^k}{dz^k}\Lambda_n(zV) - \frac{d^k}{dz^k}\Lambda(zV)\right| \to 0.$$

Applying the above estimates to $k = 2$, we get

$$\begin{aligned}\frac{d^2}{dt^2}\Lambda_n(tV)|_{t=0} &= \frac{1}{n}\mathbb{E}^\mu\left(\sum_{k=0}^{n-1}\frac{V(X_k)+V(X_{k+1})}{2} - n\mu(V)\right)^2\\ &\to Var_{\mathbb{P}_\mu}(V(X_0)) + 2\sum_{n=1}^{\infty}Cov_{\mathbb{P}_\mu}(V(X_0),V(X_n)) = \sigma^2(V).\end{aligned}$$

(b) All other properties of $J_V(r) = \sup_{t\in\mathbb{R}}(tr - \Lambda(t[V-\mu(V)]))$ are easy consequences of (4.16) and part (a) by elementary convex analysis. □

4.4 Proof of Theorem 4.1

The desired LDP of f_n^* in $(L^1, \sigma(L^1, L^\infty))$ is equivalent to the LDP of $f_n^*(x)dx$ on $M_1(\mathbb{R}^d)$ with respect to the τ-topology $\sigma(M_1(\mathbb{R}^d), b\mathcal{B})$. We divide its proof in two parts.

4.4.1 Upper bound

By the abstract Gärtner-Ellis theorem in Wu (1997, p290, Theorem 2.7) and (4.13), it is enough to show that

$$\limsup_{n\to\infty}\frac{1}{n}\log\sup_{\nu\in\mathcal{A}_{\mu,2}(L)}\mathbb{E}^\nu e^{n\int_{\mathbb{R}^d}f_n^*(y)V(y)dy} \le \Lambda(V), \tag{4.19}$$

and that $\Lambda(V)$ is monotonely continuous at 0, i.e., if (V_n) is a sequence in $b\mathcal{B}$ decreasing pointwise to 0 over $\mathbb{R}^d$, then $\Lambda(V_n) \to 0$.

The second condition is satisfied by Lemma 4.6(b). It remains to verify (4.19). Put $V_n = (K_{h_n} * V)$, then $\|V_n\| \le \|V\|$ and,

$$n\int_{\mathbb{R}^d}f_n^*(y)V(y)dy = \frac{1}{2}\sum_{k=0}^{n-1}(V_n(X_k)+V_n(X_{k+1})).$$

Consequently we have for each $\nu \in \mathcal{A}_{\mu,2}(L)$,

$$\mathbb{E}^\nu e^{n\int_{\mathbb{R}^d} f_n^*(y)V(y)dy} \le \|\frac{d\nu}{d\mu}\|_{L^2(\mu)} \cdot \|(P^{V_n})^n\|_{L^2(\mu)} \le L \cdot e^{n\Lambda(V_n)}.$$

Thus

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{E}^\nu e^{n\int_{\mathbb{R}^d} f_n^*(y)V(y)dy} \le \lim_{n\to\infty} \Lambda(V_n) = \Lambda(V)$$

where the last inequality follows from Lemma 4.6(b), for $V_n \to V,\ dx - a.e.$ So (4.19) holds.

Remark: From the upper bound above, we can derive the following exponential convergence: for any $g_1, \cdots, g_m \in b\mathcal{B}$ and for any $\delta > 0$,

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x \left(\max_{1\le i\le m} | \int_{\mathbb{R}^d} [f_n^*(x) - f(x)] g_i(x) dx| \ge \delta \right) \\
&\quad \le -\inf\{J_\mu(g);\ \max_{1\le i\le m} | \int_{\mathbb{R}^d} [g(x) - f(x)] g_i(x) dx| \ge \delta\} \\
&\quad < 0, \quad \mu - a.s.\ x \in E.
\end{aligned} \tag{4.20}$$

In fact, the last inequality follows from the inf-compactness of $J_\mu(\cdot)$ on $(M_1(E), \tau)$ and the fact that $J_\mu(\nu) = 0$ iff $\nu = \mu$ by **(A1)**. For the first inequality, let $-c(\delta)$ be the non-positive constant at the right hand side. For any $\varepsilon > 0$, using the proved upper bound, we have,

$$\int_E \sum_{n=1}^\infty \mathbb{P}_x \left(\max_{1\le i\le m} | \int_{\mathbb{R}^d} [f_n^*(x) - f(x)] g_i(x) dx| \ge \delta \right) e^{-[c(\delta)+\varepsilon]n} \mu(dx) < +\infty$$

which yields $\sum_{n=1}^\infty \mathbb{P}_x \left(\max_{1\le i\le m} | \int_{\mathbb{R}^d} [f_n^*(x) - f(x)] g_i(x) dx| \ge \delta \right) e^{-(c(\delta)+\varepsilon))n} < +\infty$, $\mu-$ $a.s.\ x$. Thus (4.20) holds $\mu - a.s.$ (for $\varepsilon > 0$ is arbitrary).

4.4.2 Lower bound

For the desired uniform lower bound, it is enough to prove that for any τ-neighborhood $N(\nu, \delta) := \{\nu' \in M_1(\mathbb{R}^d);\ |(\nu' - \nu)(g_i)| < \delta,\ i = 1, \cdots, m\}$, $g_i \in b\mathcal{B}$ with $|g_i| \le 1$, $\delta > 0$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x(f_n^*(y)dy \in N(\nu, \delta)) \ge -J_\mu(\nu),\ \mu - a.s., \tag{4.21}$$

(the arguments for this reduction, similar to those in the proof of Theorem 5.1 in Wu 2000, are left to the reader). The proof of (4.21) is divided into two steps:

Step 1. The case '$\nu = \nu_V$' for some $V \in b\mathcal{B}$. The idea of this step is borrowed from the classical works of Donsker and Varadhan (1975; 1983). Given $V \in b\mathcal{B}$, let

Q^V be the transition kernel defined in Lemma 4.5 and $\nu_V = \phi^2\mu$. By Lemma 4.5, Q^V is symmetric.

Let $\mathbb{Q}^V_{\omega(0)}$ be the law of the Markov process with transition kernel Q^V and starting point $\omega(0)$, which is $\nu_V - a.s.$ well defined on $\Omega = E^{\mathbb{N}}$, and $\mathbb{Q}^V := \int \mathbb{Q}^V_{\omega(0)} d\nu_V(\omega(0))$. Denoting by $\xi(\omega)$ the density of $\mathbb{Q}^V_{\omega(0)}$ with respect to $\mathbb{P}_{\omega(0)}$ on $\sigma(X_1)$, we have for $\mu - a.s\ \omega(0)$, on $\mathcal{F}_n := \sigma(X_k;\ 0 \le k \le n)$,

$$\left.\frac{d\mathbb{Q}^V_{\omega(0)}(d\omega_1, \cdots, d\omega_n)}{d\mathbb{P}_{\omega(0)}}\right|_{\mathcal{F}_n} = \exp\left(\sum_{k=0}^{n-1} \log \xi(\theta^k \omega)\right)$$

and $\mathbb{E}^{\mathbb{Q}^V} \log \xi = J^{(2)}(\mathbb{Q}^V|_{\mathcal{F}_1}) = J(\nu_V)$ by Lemma 4.7. For any $\varepsilon > 0$, setting

$$W_n := \{\omega : |\int_{\mathbb{R}^d} g_i(x)[f_n^*(x,\omega) - \phi^2(x)]dx < \delta,\ \forall i = 1, \cdots, m\}$$

$$D_{n,\varepsilon} := \{\omega : \frac{1}{n}\sum_{n=0}^{n-1} \log \xi(\theta^k \omega) \le J(\nu_V) + \varepsilon\},$$

we have for $\mu - a.s\ \omega(0)$,

$$\begin{aligned} \mathbb{P}_{\omega(0)}(W_n) &\ge \int_{W_n} \exp\left(-\sum_{k=0}^{n-1} \log \xi(\theta^k \omega)\right) d\mathbb{Q}^V_{\omega(0)} \\ &\ge \exp[-n(J(\nu_V) + \varepsilon)] \cdot \mathbb{Q}^V_{\omega(0)}\left(W_n \bigcap D_{n,\varepsilon}\right). \end{aligned} \tag{4.22}$$

So to get (4.21), it remains to show that $\mathbb{Q}^V_{\omega(0)}(D_{n,\varepsilon}) \to 1$ and $\mathbb{Q}^V_{\omega(0)}(W_n) \to 1$, as n goes to infinity (for any $\varepsilon > 0$), for $\mu - a.s\ \omega(0)$.

By the ergodic theorem and Fubini's theorem, we have for $\nu_V \sim \mu - a.s\ \omega(0)$,

$$\frac{1}{n}\sum_{k=0}^{n-1} \log \xi(\theta_k \omega) \to \mathbb{E}^{\mathbb{Q}^V} \log \xi = J(\nu_V), \quad \mathbb{Q}^V_{\omega(0)} - a.s.$$

which shows $\mathbb{Q}^V_{\omega(0)}(D_{n,\varepsilon}) \to 1$. To prove $\mathbb{Q}^V_{\omega(0)}(W_n) \to 1$, apply (4.20) in Remarks 4.4.1 to Q^V (instead of P) which satisfies again **(A1)** and **(A2)**, then

$$\mathbb{Q}^V_{\omega(0)}(W_n^c) \to 0,\ \nu_V - a.s.\ \omega(0).$$

The desired convergence holds.

Step 2. The general case. In order to prove (4.21) for general ν such that $J_\mu(\nu) < +\infty$, it is enough to approximate ν by ν_{V_n} as claimed in Lemma 4.8.

4.5 Proof of Theorem 4.2

The proof is divided in two parts.

Part 1. Lower bound in (4.2). The lower bound is an easy consequence of Theorem 4.1. Actually, as $\{g \in L^1;\ \|g-f\|_1 > \delta\}$ is open in the weak topology $\sigma(L^1, L^\infty)$, by Theorem 4.1, we have for any $L \geq 1$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \inf_{\nu \in \mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - f\|_1 > \delta) \geq - \inf_{g\in L^1;\ \|g-f\|_1>\delta} J(g) = -I(\delta).$$

Part 2. Upper bound in (4.2). The proof of the upper bound is much more difficult, and it is divided into three steps, where the first two steps are similar to Devroye (1983) and the third one is inspired by Louani (2000).

Step 1 (Approximation of K) As in Lei and Wu (2005), we may approximate K by $K^{(\varepsilon)} = \sum_{j=1}^m \lambda_j \dfrac{1_{A_j}}{|A_j|}$, where $\sum_{j=1}^m \lambda_j = 1$, $A_j, j \geq 1$ are some disjoint finite rectangles in $\mathbb{R}^d$ of the form $\prod_{i=1}^d [x_i, x_i + a_i)$, so it is enough to establish (4.2) only for $K = \frac{1}{|A|} 1_A$ where $A := \prod_{i=1}^d [x_i, x_i + a_i)$ (for details, see Lei and Wu 2005, step 1, Part 2, proof of Theorem 2.3). Here $|A|$ denotes the Lebesgue measure of A.

Step 2. (the method of partition) Fix such a rectangle $A := \prod_{i=1}^d [x_i, x_i + a_i)$ and $K = \frac{1}{|A|} 1_A$, and let $0 < \varepsilon < \delta/4$ be arbitrary. Since $K_{h_n} * f \to f$ in L^1, then it is enough to show that

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu \in \mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu \left(\|f_n^* - K_{h_n} * f\|_1 > \delta\right) \leq -I(\delta-). \tag{4.23}$$

Note that

$$\begin{aligned}\int |f_n^*(x) - K_{h_n} * f(x)| dx &\leq \int \left| \frac{1}{|A| h_n^d} \int_{x+hA} L_n(dy) - \frac{1}{|A| h_n^d} \int_{x+h_nA} f(y) dy \right| dx \\ &\leq \frac{1}{|A| h_n^d} \int |L_n(x + h_n A) - \mu(x + h_n A)| dx.\end{aligned}$$

Consider the partition of $\mathbb{R}^d$ into sets B that are d-fold products of intervals of the form $[\dfrac{(i-1)h_n}{p}, \dfrac{ih_n}{p})$, where $i \in \mathbb{Z}$, and $p \in \mathbb{N}^*$ such that $\min_i a_i \geq \dfrac{2}{p}$. Call the partition Ψ.

Let $A^* = \Pi_{i=1}^d [x_i + \dfrac{1}{p}, x_i + a_i - \dfrac{1}{p})$. We have

$$C_x := (x + h_n A) \backslash \bigcup_{B \in \Psi, B \subseteq x + h_n A} B \subseteq x + h_n (A \backslash A^*).$$

Consequently,

$$\begin{aligned}&\int |f_n^*(x) - K_{h_n} * f(x)|dx \\ &\le \frac{1}{|A|h_n^d}\int \sum_{B\in\Psi, B\subseteq x+hA} |L_n(B)-\mu(B)|dx + \frac{1}{|A|h_n^d}\int \{\mu(C_x)+L_n(C_x)\}dx.\end{aligned} \tag{4.24}$$

Using the fact that for any set $C \in \mathcal{B}$, $h > 0$ and any probability measure ν on $\mathbb{R}^d$,

$$\int \nu(x+hC)dx = |hC| = h^d|C|$$

(using Fubini's theorem), the last term in (4.24) is bounded from above by

$$\frac{1}{|A|h_n^d} 2h_n^d |A\backslash A^*| = 2\left(1 - \prod_{i=1}^d \left(1 - \frac{2}{pa_i}\right)\right) \le \varepsilon$$

when p is large enough. We fix such p which is independent of n.

For any finite constant $R > 0$, letting $S_{OR} := \{x \in \mathbb{R}^d;\ |x| \le R\}$, we can bound the first term in the righ hand side of (4.24) from above by

$$\begin{aligned}&\sum_{B\in\Psi, B\cap S_{OR}\neq\emptyset} |L_n(B)-\mu(B)| \frac{1}{|A|h_n^d}\int_{B\subseteq x+h_nA} dx \\ &\qquad + \frac{1}{|A|h_n^d}\int_{B\subseteq x+h_nA} dx \{L_n(S_{OR}^c) - \mu(S_{OR}^c) + 2\mu(S_{OR}^c)\}.\end{aligned}$$

Clearly, $h_n^{-d}\int_{B\subseteq x+h_nA} dx \le |A|$, and $\mu(S_{OR}^c) < \varepsilon/2$ for $R \ge R_0$ large enough.

By Lemma 4.5, we have for all $t > 0$,

$$\begin{aligned}&\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu\{L_n(S_{OR}^c) - \mu(S_{OR}^c) > \varepsilon\} \\ &\qquad \le -J_{1_{S_{OR}^c}}(\varepsilon) \le -\left(t[\varepsilon + \mu(S_{OR}^c)] - \Lambda(t1_{S_{OR}^c})\right).\end{aligned}$$

Since $\lim_{R\to\infty}\Lambda(t1_{S_{OR}^c}) = 0$ by Lemma 4.6, for any $M > 0$, then the left hand side above is bounded from above by $-M$ for all R large enough, say $R \ge R_1$. Fix such $R \ge R_0 \vee R_1$ below. Summarizing those estimations, we obtain

$$\begin{aligned}&\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu\left(\int |f_n^*(x) - K_{h_n} * f(x)|dx > \delta\right) \\ &\quad \le (-M) \vee \limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu\left(\sum_{B\in\Psi, B\cap S_{OR}\neq\emptyset} |L_n(B)-\mu(B)| > \delta - 3\varepsilon\right).\end{aligned} \tag{4.25}$$

Step 3. It remains to control the last term in (4.25). Set

$$\tilde{\Psi} = \{B; B \in \Psi, B \cap S_{OR} \neq \emptyset\},\ C := \left(\bigcup_{B \in \tilde{\Psi}} B\right)^c$$

and $\mathcal{B}(\tilde{\Psi}) = \sigma\{B; B \in \tilde{\Psi}\}$, the σ-field generated by $\tilde{\Psi}$. Regarding L_n and μ as probability measures on $\mathcal{B}(\tilde{\Psi})$, and denoting the total variation of $L_n - \mu$ on $\mathcal{B}(\tilde{\Psi})$ by $\|L_n - \mu\|_{\mathcal{B}(\tilde{\Psi})}$, we have

$$\sum_{B \in \Psi, B \cap S_{OR} \neq \emptyset} |L_n(B) - \mu(B)| \leq \|L_n - \mu\|_{\mathcal{B}(\tilde{\Psi})} = \max_{V \in \{-1,1\}^{\tilde{\Psi}}} \left(L_n(V) - \mu(V)\right),$$

where $\{-1,1\}^{\tilde{\Psi}}$ denotes the set of all $\mathcal{B}(\tilde{\Psi})$-measurable functions with values in $\{-1,1\}$ (which can be identified as the set of functions from $\tilde{\Psi}$ to {-1,1}) . Therefore, for any fixed $r > 0$,

$$\mathbb{P}_\nu \left(\sum_{B \in \Psi, B \cap S_O R \neq \emptyset} |L_n(B) - \mu(B)| > r \right) \leq \mathbb{P}_\nu \left(\max_{V \in \{-1,1\}^{\tilde{\Psi}}} L_n(V) - \mu(V) > r \right)$$
$$\leq \sum_{V \in \{-1,1\}^{\tilde{\Psi}}} \mathbb{P}_\nu \left(L_n(V) - \mu(V) > r\right).$$

By Lemma 4.5(b), for each $V \in \{-1,1\}^{\tilde{\Psi}}$ and for all $r > 0$,

$$\sup_{\nu \in \mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu \left(L_n(V) - \mu(V) > r\right) \leq \sup_{\nu \in \mathcal{A}_{\mu,2}(L)} \exp\left(-nJ_V(r)\right) \left\|\frac{d\nu}{d\mu}\right\|_2 \leq L \exp\left(-nJ_V(r)\right).$$

Secondly, the number of elements $\tilde{\Psi}$ is no greater than $\left(\frac{2Rp}{h_n} + 2\right)^d + 1 = o(n)$ by (4.3), and $\{-1,1\}^{\tilde{\Psi}}$ has $2^{\#\tilde{\Psi}} = 2^{o(n)}$ elements for n large enough. Consequently letting $\mathbb{B}(1)$ be the unit ball in $L^\infty(\mu)$, we have

$$\limsup_{n \to \infty} \frac{1}{n} \log \sup_{\nu \in \mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu \left(\sum_{B \in \Psi, B \cap S_O R \neq \emptyset} |L_n(B) - \mu(B)| > r \right)$$
$$\leq \limsup_{n \to \infty} \frac{1}{n} \log 2^{o(n)} L \sup_{V \in \mathbb{B}(1)} \exp\left(-nJ_V(r)\right) = -\inf_{V \in \mathbb{B}(1)} J_V(r).$$

Combining with (4.25), we get

$$\limsup_{n \to \infty} \frac{1}{n} \log \sup_{\nu \in \mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu \left(\int |f_n^*(x) - K_{h_n} * f(x)| dx > \delta \right)$$
$$\leq (-M) \vee \left(-\inf_{V \in \mathbb{B}(1)} J_V(\delta - 3\varepsilon) \right).$$

Since J_V is convex, nondecreasing and left continuous on $[0,+\infty)$, consequently using $\|\nu-\mu\|_{TV} = \sup_{\|V\|\le 1}[\nu(V)-\mu(V)] = 2\sup_{A\in\mathcal{B}}|\nu(A)-\mu(A)|$ and (4.15), we have

$$\begin{aligned} I(\delta) &= \inf\{J_\mu(\nu) | \sup_{\|V\|\le 1}[\nu(V)-\mu(V)] > \delta\} \\ &= \inf_{\|V\|\le 1}\inf_{r>\delta} J_V(r) = \inf_{\|V\|\le 1} J_V(\delta+) \end{aligned} \tag{4.26}$$

As $M>0$ are arbitrary and $\lim_{\varepsilon\to 0+}\inf_{V\in\mathbb{B}(1)} J_V(\delta-3\varepsilon) = I(\delta-)$ by (4.26), we obtain the desired (4.23) and then complete the proof of the upper bound in (4.2).

4.6 Proof of Theorem 4.3

The proof is divided in two parts, where the first part is for the upper bound and the second is for the lower bound.

Part 1. Large deviation upper bound. This is an easy consequence of Theorem 4.1. In fact, for any $g\in L^1$ and δ fixed, as $\{\tilde g\in L^1;\ \|\tilde g-g\|_1\le\delta\}$ is closed in the weak topology $\sigma(L^1,L^\infty)$, then by Theorem 4.1,

$$\limsup_{n\to\infty}\frac{1}{n}\log\sup_{\nu\in\mathcal{A}_{\mu,2}(L)}\mathbb{P}_\nu(\|f_n^*-g\|_{L^1(\mathbb{R}^d)}<\delta)\le -\inf_{\tilde g;\ \|\tilde g-g\|_1\le\delta}J(\tilde g).$$

Letting $\delta\to 0$, we get the desired result by the lower semi-continuity of J (which follows from (4.13)).

Part 2. Large deviation lower bound. It is enough to prove that $\forall g\in\mathcal{P}(E)$,

$$\lim_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}_x(\|f_n^*-g\|_{L^1(\mathbb{R}^d)}<\delta) = -J(g),\ \ \mu-a.s.$$

(This implies the desired uniform lower bound as in Wu (2000).) The proof is divided in two steps.

Step 1. For '$gdx=\nu_V$' case. The proof of this case is parallel to the one of Step 1 in the proof of (4.21): the only change is that we now set

$$W_n := \{\omega : \|f_n^*(\omega)-g\|_1<\delta\}$$

and the key point is to prove $\mathbb{Q}^V_{\omega(0)}(W_n)\to 1,\ \nu_V-a.s.$. Applying the upper bound in Theorem 4.2 to Q^V (instead of P), we have $\mathbb{Q}^V(W_n^c)\to 0$ at exponential rate. Using Borel-Cantelli's lemma, $\mathbb{Q}^V_{\omega(0)}(W_n^c)\to 0$ for $\nu_V-a.s.\ \omega(0)$.

Step 2. The general case. For completing the proof, it remains to show the claim below:

$\forall \nu = gdx \in M_1(\mathbb{R}^d)$ satisfies $J(g) < +\infty$, there exists a sequence of (ν_{V_n}), such that $\|\nu_{V_n} - \nu\|_{TV} \to 0$, and $\limsup_{n\to\infty} J(\nu_{V_n}) \le J_\mu(\nu)$.

This has been settled in Lemma 4.8.

4.7 Proof of Theorem 4.4

We begin with

Lemma 4.10 *Given $V \in b\mathcal{B}$. If T_n is an asymptotically consistent estimator of $\langle V, f\rangle := \int_E V(x)f(x)dx$, i.e., for each $(P,\mu) \in \Theta$ (satisfying* **(A1)** *and* **(A2)***), $|\langle T_n, V\rangle - \langle f, V\rangle| \to 0$ in probability $\mathbb{P}_\mu$, then*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu \left(\langle T_n - f, V\rangle > \delta\right) \ge -\inf\{J(g);\ \langle g - f, V\rangle > \delta\}. \tag{4.27}$$

Proof. It is enough to prove that the left hand side of (4.27) is greater than $-J(g)$ for every $g \in \mathcal{P}$ which satisfies $\langle g - f, V\rangle > \delta$ and $J(g) < +\infty$. By Step 2 in the proof of the lower bound of Theorem 4.1, it suffices to prove it in the case of $g = \nu_{\tilde V}$ where $\tilde V \in b\mathcal{B}$ is arbitrary. The proof, completely parallel to Step 1 in the proof of the lower bound of Theorem 4.1, is based on the fact that $(Q^{\tilde V}, \nu_{\tilde V}) \in \Theta$ again, so it is omitted. □

Lemma 4.11 *Under* **(A1)** *and* **(A2)***, let $I(\cdot)$ be defined as in Theorem 4.2 (4.8). Then*

$$\lim_{r\to 0+} \frac{I(r)}{r^2} = \frac{1}{2\sup_{\|V\|\le 1} \sigma^2(V)} = \frac{1}{8\sup_{A\in\mathcal{B}(E)} \sigma^2(1_A)}. \tag{4.28}$$

Proof. We only prove the first equality in (4.28) (the proof of the second one is easy). By (4.26) and Lemma 4.9(b), for any $V \in b\mathcal{B}$ with $\|V\| \le 1$,

$$\limsup_{r\to 0} \frac{I(r)}{r^2} \le \lim_{r\to 0} \frac{J_V(r+)}{r^2} = \frac{1}{2\sigma^2(V)},$$

so $\lim_{r\to 0+} \dfrac{I(r)}{r^2} \le \dfrac{1}{2\sup_{\|V\|\le 1} \sigma^2(V)}$.

For the converse inequality, let $L > 1$ be arbitrary but fixed. For any $\delta > 0$ small enough, we have by Lemma 4.9,

$$C(L\delta) := \sup_{t\in[0,L\delta]} \sup_{V\in\mathbb{B}(1)} \left|\frac{d^3}{dt^3}\Lambda(tV)\right| < +\infty.$$

Thus by Taylor's formula at order 3, we get for any $V \in \mathbb{B}(1)$ and $r \in (0, \delta]$,

$$
\begin{aligned}
J_V(r) &\geq \sup_{t\in[0,Lr]} (tr - \Lambda(t[V - \mu(V)])) \geq \sup_{t\in[0,Lr]} \left(tr - \frac{t^2\sigma^2(V)}{2} \right) - \frac{(Lr)^3}{6} \cdot C(L\delta) \\
&\geq r^2 \left(L \wedge \sigma^{-2}(V) - \frac{[L \wedge \sigma^{-2}(V)]^2\sigma^2(V)}{2} \right) - \frac{(Lr)^3}{6} \cdot C(L\delta)
\end{aligned}
$$

where the last inequality is obtained by taking $t = r[L \wedge \sigma^{-2}(V)]$. Thus by (4.26),

$$
\begin{aligned}
\liminf_{r\to 0+} \frac{I(r)}{r^2} &= \liminf_{r\to 0+} \inf_{V\in\mathbb{B}(1)} \frac{J_V(r)}{r^2} \\
&\geq \min \left\{ \inf_{V\in\mathbb{B}(1):\ \sigma^{-2}(V)\leq L} \frac{1}{2\sigma^2(V))};\ \inf_{V\in\mathbb{B}(1):\ \sigma^{-2}(V)>L} (L - L/2) \right\} \\
&\geq \min \left\{ \inf_{V\in\mathbb{B}(1)} \frac{1}{2\sigma^2(V)};\ \frac{L}{2} \right\}
\end{aligned}
$$

where the desired converse inequality follows from letting $L \to +\infty$. □

Proof. [Proof of Theorem 4.4] (a) By Lemma 4.10, since $\mathcal{D}$ is dense in the unit ball of L^∞ with respect to $\sigma(L^\infty, L^1)$,

$$
\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\|T_n - f\|_1 > r) &\geq \sup_{V\in\mathcal{D}} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\langle T_n - f, V\rangle > r) \\
&\geq - \inf_{V\in\mathcal{D}} \inf\{J(g) | \langle g - f, V\rangle > r\} \\
&= - \inf\{J(g) | \sup_{V\in\mathcal{D}} \langle g - f, V\rangle > r\} = - \inf_{g:\|g-f\|_1>r} J(g) = -I(r).
\end{aligned}
$$

Thus (4.10) follows from Lemma 4.11. The second claim easily follows from (4.10) by means of the extra condition on T_n and **(A1)**.

(b) It follows from Theorem 4.3 and Lemma 4.11. □

Acknowledgments: I am very grateful to Professor L. Wu for his advice and guidance and to the referees for their careful comments and constructive suggestions.

Bibliography

[1] Bosq, D., Merlevède, F. and Peligrad, M. (1999) Asymptotic normality for density kernel estimators in discrete and continuous time. *J. Multivariate Anal.*, **68**, 78-95.

[2] Csörgö, M. and Horváth, L. (1988) Central limit theorems for L_p-norms of density estimators. *Z. Wahrsch. Verw. Gebiete.*, **80**, 269-291.

[3] Dembo, A. and Zeitouni, O. (1998) *Large Deviations Techniques and Applications.* 2nd Edn, Springer, New York.

[4] Deuschel, J.D. and Stroock, D.W. (1989) *Large Deviations.* Pure Appl. Math., vol 137, Academic Press, San Diego.

[5] Devroye, L. (1983) The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates. *Ann. Statist.*, **11**, 896-904.

[6] Donsker, M.D. and Varadhan, S.R.S. (1975; 1983) Asymptotic evaluation of certain Markov process expectations for large time, I-IV. *Commun. Pure Appl. Math.*, (1975) **28**, 1-47, 279-301; **29** 389-461; (1983) **36** , 183-212.

[7] Gao, F.Q., (2003) Moderate deviations and large deviations for kernel density estimators. *J. Theoret. Probab.*, **16**, 401-418.

[8] Gao, F.Q., Moderate deviations and the law of iterated logarithm for kernel density estimators in L_1. Preprint 2003.

[9] Giné, E., Mason, D.M. and Zaitsev, A.Yu. (2003) The L_1-norm density estimator process. *Ann. Probab.*, **31**, 719-768.

[10] Gong, F.Z. and Wu, L.M., (2000) spectral gap of positive operators and applications. *C. R. Acad. Sci. Paris Sér. I Math.*, **331**, 983-988.

[11] Horváth, L. (1991) On L_p-norms of multivariate density estimators. *Ann. Statist.*, **19**, 1933-1949.

[12] Kato, T. (1992) *Perturbation Theory for Linear Operators.* 2nd Edn, Spring-Berlin, New York/Berlin.

[13] Lei, L.Z., Wu, L.M. and Xie, B. (2003) Large deviations and deviation inequality for kernel density estimator in $L_1(R^d)$-distance. In *Development of Modern Statistics and Related Topics*, Ser. Biostat. 1., 89-97, *World Sci. Publishing, River Edge, NJ.*

[14] Lei, L.Z. and Wu, L.M. (2004) The exponential convergence of kernel density estimator in L_1 for ϕ-mixing processes. *Ann.I.S.U.P.*, **48**, 59-68.

[15] Lei, L.Z. and Wu, L.M. (2005) Large deviations of kernel density estimator in L_1 for uniformly ergodic Markov processes. *Stochastic Process. Appl.*, **115**, 275-298.

[16] Louani, D. (2000) Large deviations for the L_1-distance in kernel density estimation. *J. Statist. Plann. Inference.*, 177-182.

[17] Meyn, S.P. and Tweedie, R.L. (1993) *Markov chains and stochastic stability.* Springer, Berlin.

[18] Peligrad, M. (1992) Properties of uniform consistency of the kernel estimators of density and of regression functions under dependence assumption. *Stochastics Stochastics Rep.*, **40**, 147-168.

[19] Wu, L.M. (1995) Moderate deviations of dependent random variables related to CLT. *Ann. Probab.*, **23**, 420-445.

[20] Wu, L.M. (1997) An introduction to large deviations. In: Yan, J.A., Peng, S.G., Fang, S.Z. and Wu, L.M. (Eds), *Several Topics in Stochastic Analysis*, 225-336, Academic Press of China, Beijing, (In Chinese).

[21] Wu, L.M. (2000) Uniformly integrable operator and large deviations for Markov processes. *J. Funct. Anal.*, **172**, 301-376.

[22] Wu, L.M. (2000b) A deviation inequality for non-reversible Markov processes. *Ann. Inst. H. Poincaré Probab. Statist.*, **36**, 435-445.

[23] Wu, L.M. (2002) Large deviations for reversible Markov chains. Preprint.

Chapter 5

Strong Consistency and CLT for the Random Decrement Estimator

This article is a joint work with Professor Pierre Bernard.

5.1 Introduction

In the goal of maintenance and damage detection of big structures (like cable stayed bridges), the use of controlled excitations is not possible. For at least two reasons: this should be too difficult to realize, and the structure should be set out of service for some time. Hence techniques using the dynamical response under ambient loadings (like wind and traffic) are of very high interest.

This is the case of the Random Decrement algorithm, introduced in the late sixties by H.A.Cole, a NASA engineer. This technique, which was empirically designed, is simple to use, very performing, and appears to be robust with respect of the nature of loading.

The first tentative of analysis was made by Caughey [3] in 1961, with the following arguments.The stationary response of a White Noise excited linear structure is the sum of two terms: one term depending on the initial conditions, and one random term, assumed to have zero mean, which is the convolution product of the excitation by the impulse response function of the structure. By a statistical averaging of pieces of trajectories beginning with the same initial conditions, the random part should cancel. The first part only is maintained, and this part is the free response of the system to some particular initial conditions. So, identification methods designed to determine characteristics of the structure from the data of the free response to an impulse (as for example Ibrahim method [8]) can be used.

This idea is good, but as we will prove, the result is not quite as expected from this heuristic approach. The mathematical analysis is somewhat more complicated, and gave rise to many errors in the engineering literature on this question. However, this method appeared to be very performing, once the expected result was clearly established.

In a more general setting, applied to stationary ergodic Gaussian processes, the random decrement method furnishes a powerful estimator of its covariance function, which can be biased, which is probably numerically the more efficient estimator from our experience. Mechanicians use to compare the power of this algorithm to that of the FFT algorithm for discrete time Fourier Transform.

5.2 Setting of the problem

Let $\{X_i, i \geq 0\}$ be a scalar stationary ergodic Gaussian sequence with zero mean and variance σ^2 where $\sigma > 0$. Denote by Φ the distribution function of the canonical Gaussian measure on the real line, by $\rho(j) = \dfrac{Cov(X_j, X_0)}{\sqrt{Var(X_j)Var(X_0)}}$ the correlation coefficient between X_j and X_0 for any relative integer j.

Introduce some condition known as **triggering condition** in the random decrement approach:

$$\mathcal{D}_k = (X_k, X_{k+1}, \ldots, X_{k+d-1}) \in \Delta; \tag{5.1}$$

where d is a fixed integer and Δ is a convenient domain in $\mathbb{R}^d$. The main applications in mechanics (level crossings or upcrossings) deals respectively with $d = 1$ and $d = 2$. The meaning of such condition has to be properly interpreted: condition (5.1) defines in fact a time valued point process, as those (random) times for which the condition is satisfied. Denote by $(\tau_k, k \in \mathbb{Z})$ this stationary time process. This process will be defined for $k > 0$ as follows:

$$\begin{aligned} \tau_1 &= \inf\{j \geq 0 : (X_j, X_{j+1}, \ldots, X_{j+d-1}) \in \Delta\}, \\ \tau_{k+1} &= inf\{j > \tau_k : (X_j, X_{j+1}, \ldots, X_{j+d-1}) \in \Delta\}; \end{aligned} \tag{5.2}$$

For $k \leq 0$, it will be defined as the stationary extension of this sequence. Let $l(n)$ be its counting function, that is the number of τ_k's in the set $\{0, 1, \ldots, n-1\}$.

Set

$$D_n(j) := \frac{1}{n} \sum_{k=1}^{n} X_{\tau_k + j}, \tag{5.3}$$

and

$$\bar{D}_n(j) := \frac{1}{l(n)} \sum_{k=1}^{l(n)} X_{\tau_k + j}.$$

In this paper, we establish a Law of Large Numbers (LLN) and a Central Limit Theorem (CLT) for $D_n(j), n > 0$ and $\bar{D}_n(j), n > 0$, for any fixed j. The limit will be expressed in terms of some conditional expectation, which will be calculated in some important cases for applications. Moreover, in the particular case where the stationary sequence is obtained by discretization with step h from a continuous time process, the limit as $h \to 0$ will be calculated. This gives useful approximations for small h, and enlightens the problem of conditioning in this case.

We denote by $\mathcal{F}_a^b$ the σ-field generated by the random variables $(X_i)_{a \leq i \leq b}$, $-\infty \leq a \leq b \leq \infty$, and by $L^2(\mathcal{F}_a^b)$ the set of all $\mathcal{F}_a^b$-measurable random variables with finite variance. The maximal correlation coefficient $\rho^*(n)$ between the past $\{X_i, i \leq 0\}$ and the future $\{X_i, i \geq n\}, n > 0$, of the stationary sequence $(X_i, i \in \mathbb{Z})$ is defined by

$$\rho^*(n) = \sup_{f \in L^2(\mathcal{F}_{-\infty}^0),\ g \in L^2(\mathcal{F}_n^\infty)} \frac{Cov(f,g)}{\sqrt{Var(f)Var(g)}}.$$

Furthermore,

$$\rho^*(n) = \sup_{f \in L2(\mathcal{F}_n^\infty), \mathbb{E}f=0} \frac{\|\mathbb{E}(f|\mathcal{F}_0)\|_2}{\|f\|_2}.$$

A very beautiful result of Sarmanov [12] says that for stationary Gaussian processes,

$$\rho^*(n) = \sup_{\xi,\eta} \rho(\xi,\eta) \tag{5.4}$$

where the supremum is taken over all (ξ,η) such that $\xi = \sum_{i \leq 0} a_i X_i$ and $\eta = \sum_{j \geq n} b_j X_j$ (finite linear combinations).

We shall compare it to the strong condition. The process $\{X_i\}$ is said to satisfy the strong mixing condition if

$$\sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_n^\infty} |P(A \cap B) - P(A)P(B)| = \alpha(n) \downarrow 0,\ n \to \infty.$$

In the particular case of Gaussian distribution, it has been proved by Kolmogorov and Rozanov [9] that $\alpha(n) \leq \rho^*(n) \leq 2\pi\alpha(n)$.

5.3 Main results

5.3.1 Discrete time case

Our first result is about the (strong) Law of Large Number for the estimator $D_n(j)$ and $\bar{D}_n(j)$.

Theorem 5.1 *$\{X_i, i \geq 0\}$ is a stationary ergodic Gaussian process with the marginal normal distribution $\mathcal{N}(0,\sigma^2)$, then,*

$$\lim_{n\to\infty} D_n(j) = \lim_{n\to\infty} \bar{D}_n(j) == \mathbb{E}(X_j|\Delta),\ a.s.$$

In statistics, the CLT plays a central role. This is the purpose of

Theorem 5.2 *$\{X_i, -\infty \le i \le \infty\}$ is a stationary ergodic Gaussian process with the marginal normal distribution $\mathcal{N}(0, \sigma^2)$. If*

$$\sum_{n=0}^{\infty} \rho^*(n) < +\infty, \tag{5.5}$$

then,

$$\sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} X_{\tau_k+j} - \mathbb{E}(X_j|\Delta)\right) \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}^2)$$

where

$$\begin{aligned}\tilde{\sigma}^2 = \frac{1}{P(\Delta_0)} & \left[Var\left(1_{\Delta_0}\left(X_j - \mathbb{E}(X_j|\Delta_0)\right)\right)\right. \\ & \left. +2\sum_{k=1}^{+\infty} Cov\left(1_{\Delta_0}\left(X_j - \mathbb{E}(X_j|\Delta_0)\right), 1_{\Delta_k}\left(X_{k+j} - \mathbb{E}(X_{k+j}|\Delta_k)\right)\right)\right].\end{aligned} \tag{5.6}$$

with $\Delta_k = \{\omega; (X_k, \cdots, X_{k+d-1}) \in \Delta\}$, and the last series is absolutely convergent.

Now let us study a particular case: $d = 2$ and for $a \ge 0$,

$$\Delta(+, a) = \{(x, y) \in \mathbb{R}^2; x \le a, y \ge a\}.$$

In that case, $\sum_{k=0}^{n-1} 1_{\Delta_k}$ is the up-crossing number of the level $a \ge 0$ during time interval $[0, n]$. The following explicit result will be crucial for the computation of the asymptotic bias of the RDE in the continuous time case.

Proposition 5.3 *For a stationary Gaussian process (X_n), let $\rho_n = \rho(X_0, X_n)$ and $\Delta = \Delta(+, a)$. Then*

$$\mathbb{E}(X_j|(X_0, X_1) \in \Delta) = \frac{C_0 f(a, \rho_1) + C_1(-f(-a, \rho_1))}{g(a, \rho_1)},$$

where

$$C_0 = \frac{\rho_j - \rho_1\rho_{j-1}}{1 - \rho_1^2}, \quad C_1 = \frac{\rho_{j-1} - \rho_1\rho_j}{1 - \rho_1^2}$$

and

$$f(a,\rho_1) = -\frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2\sigma^2}\right) + (1+\rho_1)\frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2\sigma^2}\right) \Phi\left(\frac{a-\rho_1 a}{\sigma\sqrt{1-\rho_1^2}}\right),$$

$$g(a,\rho_1) = \int_{\rho_1}^{1} u(\rho)d\rho,\ u(\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{a^2}{(1+\rho)\sigma^2}\right).$$

In particular, when $a=0$,

$$\mathbb{E}(X_j|(X_0,X_1)\in\Delta) = \frac{(\rho_{j-1}-\rho_j)\sigma}{\sqrt{\frac{2}{\pi}(\frac{\pi}{2} - \arcsin\rho_1)}}.$$

5.3.2 Continuous time case

Now let $(X_t)_{t\geq 0}$ be a centered stationary ergodic Gaussian process such that $\rho(t) := \rho(X_0, X_t)$ is continuous. It is extended to $\mathbb{R}$ by $\rho(-t) := \rho(t)$ for all $t \geq 0$. By Bochner's theorem, there exists a positive bf symmetric bounded measure μ on $\mathbb{R}$ such that

$$\rho(t) = \sigma^{-2}\int_{\mathbb{R}} e^{itx} d\mu(x),\ \forall t\in\mathbb{R}.$$

μ is the so called spectral measure of (X_t).

Since $\mathbb{E}(X_t|X_0 = a) = a\rho(t)$, the RDE for estimating $a\rho(t)$ with $t > 0$ is defined by

$$D_n(h,t) = \frac{1}{n}\sum_{k=1}^{n} X_{\tau_k(h)+t}$$

where $\tau_k(h)$ is the successive times of (X_{kh}) for upcrossing the level a, and the mesh h is small so that $t/h \in \mathbb{N}^*$. By Theorem 5.1, we have with probability one,

$$\lim_{n\to\infty} D_n(h,t) = \mathbb{E}\left(X_t | X_0 \leq a, X_h \geq a\right).$$

Theorem 5.4 *Let $(X_t)_{t\geq 0}$ be a centered stationary ergodic Gaussian process such that $\rho(t) := \rho(X_0, X_t)$ is continuous on $\mathbb{R}^+$ and derivable for $t > 0$.*

(a) If

$$\Gamma = \int_{\mathbb{R}} x^2 d\mu(x) < +\infty,\ (\textit{equivalently}\quad \rho''(0) > -\infty)$$

then we have with probability one,

$$\lim_{h=T/m\to 0+}\lim_{n\to\infty} D_n(h,t) = \rho(t)a - \sqrt{\frac{\pi}{2\Gamma}}\rho'(t)\sigma^2 = \rho(t)a - \sqrt{\frac{\pi}{2}}\frac{\rho'(t)\sigma}{\sqrt{-\rho''(0)}}$$

(b) If $\Gamma = \int_{\mathbb{R}} x^2\mu(dx) = +\infty$*, then we have a.s.*

$$\lim_{h=T/m\to 0+}\lim_{n\to\infty} D_n(h,t) = \rho(t)a.$$

Furthermore, both parts (a) and (b) hold true for $\bar{D}_n(h,t)$*.*

By Chung [4, Theorem 6.4.1], if μ has a finite absolute moment of order 2 (i.e., $\Gamma < +\infty$), then $\rho(t)$ has a continuous derivative of order 2 given by $\rho''(t) = -\sigma^{-2}\int e^{itx}x^2 d\mu(x)$. Conversely, if $\rho(t)$ has a finite derivative of order 2 at $t = 0$, then μ has a finite moment of order 2.

Note also that $1/\sigma^2\mu(dx)$ is the distribution of the noise frequence. So this theorem just says that $D_n(h,t), \bar{D}_n(h,t)$ are biased if the process contains sufficiently numerous high frequences of noise.

5.4 Several exact calculus

The following lemma is elementary.

Lemma 5.5 *Let* $\{X_i, i \geq 0\}$ *be a stationary ergodic Gaussian process with the standard distribution* $\mathcal{N}(0,\sigma^2)$*, then*

$$\mathbb{E}(X_j|X_0,X_1) = C_0X_0 + C_1X_1,$$

where

$$C_0 = \frac{\rho_j - \rho_1\rho_{j-1}}{1-\rho_1^2}, \quad C_1 = \frac{\rho_{j-1} - \rho_1\rho_j}{1-\rho_1^2}.$$

Proof. [Proof of Proposition 5.3]

At first we have by Lemma 5.5,

$$\begin{aligned}\mathbb{E}(X_j|(X_0,X_1)\in\Delta) &= \frac{\mathbb{E}(X_j 1_{(X_0,X_1)\in\Delta})}{\mathbb{P}((X_0,X_1)\in\Delta)} \\ &= \frac{C_0\mathbb{E}(X_0 1_{(X_0,X_1)\in\Delta}) + C_1\mathbb{E}(X_1 1_{(X_0,X_1)\in\Delta})}{\mathbb{E}(1_{(X_0,X_1)\in\Delta})}\end{aligned} \tag{5.7}$$

where $\Delta = \Delta(+,a) = \{(x,y); x \leq a, y \geq a\}$. And put

$$\begin{aligned} f(a,\rho_1) &:= \mathbb{E}(X_0 1_{(X_0,X_1)\in\Delta}) \\ g(a,\rho_1) &:= \mathbb{E}(1_{(X_0,X_1)\in\Delta}). \end{aligned}$$

The calculus will be divided to several steps for clarity.

Step 1: $f(a,\rho_1) = \mathbb{E}(X_0 1_{(X_0,X_1)\in\Delta})$

To calculate $f(a,\rho_1)$, we take the conditional expectation w.r.t X_0,

$$\begin{aligned}\mathbb{E}(X_0 1_{(X_0,X_1)\in\Delta}) &= \mathbb{E}(\mathbb{E}(X_0 1_{(X_0,X_1)\in\Delta}|X_0)) \\ &= \mathbb{E}(X_0 1_{X_0\le a}\mathbb{E}(1_{X_1>a}|X_0))\end{aligned} \tag{5.8}$$

As $X_1 = \rho_1 X_0 + \sigma\sqrt{1-\rho_1{}^2}\xi$, where $\xi \sim \mathcal{N}(0,1)$ is independent of X_0, so we have

$$\begin{aligned}\mathbb{E}(X_0 1_{X_0\le a}\mathbb{E}(1_{X_1>a}|X_0)) &= \mathbb{E}\left(X_0 1_{X_0\le a}\mathbb{P}\left(\xi > \frac{a-\rho_1 X_0}{\sigma\sqrt{1-\rho_1^2}}|X_0\right)\right) \\ &= \mathbb{E}\left[X_0 1_{X_0\le a}\Phi\left(\frac{\rho_1 X_0 - a}{\sigma\sqrt{1-\rho_1^2}}\right)\right] \\ &= \int_{-\infty}^{a}\Phi\left(\frac{\rho_1 x - a}{\sigma\sqrt{1-\rho_1^2}}\right)\frac{x}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right)dx \\ &= \int_{-\infty}^{a}\Phi(\frac{\rho_1 x - a}{\sigma\sqrt{1-\rho_1^2}})\frac{-\sigma}{\sqrt{2\pi}}d\left(\exp\left(-\frac{x^2}{2\sigma^2}\right)\right) \\ &= -\frac{\sigma}{\sqrt{2\pi}}\Phi\left(\frac{\rho_1 a - a}{\sigma\sqrt{1-\rho_1^2}}\right)\exp\left(-\frac{a^2}{2\sigma^2}\right) \\ &\quad + \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{a}\exp\left(-\frac{x^2}{2\sigma^2}\right)\frac{d}{dx}\Phi\left(\frac{\rho_1 x - a}{\sigma\sqrt{1-\rho_1^2}}\right)dx\end{aligned} \tag{5.9}$$

Noting that,

$$\begin{aligned}&\int_{-\infty}^{a}\exp\left(-\frac{x^2}{2\sigma^2}\right)\frac{d}{dx}\Phi\left(\frac{\rho_1 x - a}{\sigma\sqrt{1-\rho_1^2}}\right)dx \\ &= \int_{-\infty}^{a}\exp\left(-\frac{x^2}{2\sigma^2}\right)\frac{\rho_1}{\sigma\sqrt{1-\rho_1^2}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{\rho_1 x - a}{\sigma\sqrt{1-\rho_1^2}}\right)^2\right\}dx \\ &= \frac{\rho_1}{\sigma\sqrt{1-\rho_1^2}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{a^2}{2\sigma^2}\right)\int_{-\infty}^{a}\exp\left\{-\frac{1}{2}\left(\frac{x - a\rho_1}{\sigma\sqrt{1-\rho_1^2}}\right)^2\right\}dx \\ &= \frac{\rho_1}{\sigma\sqrt{1-\rho_1^2}}\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-a^2}{2\sigma^2}\right)\sigma\sqrt{1-\rho_1^2}\int_{-\infty}^{\frac{a-a\rho_1}{\sigma\sqrt{1-\rho_1^2}}}\exp\left(-\frac{u^2}{2}\right)du \\ &= \rho_1\exp\left(\frac{-a^2}{2\sigma^2}\right)\Phi\left(\frac{a-\rho_1 a}{\sigma\sqrt{1-\rho_1^2}}\right)\end{aligned}$$

and $\Phi(-x) = 1 - \Phi(x)$, so we obtain

$$\begin{aligned}
f(a,\rho_1) &= -\frac{\sigma}{\sqrt{2\pi}}\Phi\left(\frac{\rho_1 a - a}{\sigma\sqrt{1-\rho_1^2}}\right)\exp\left(-\frac{a^2}{2\sigma^2}\right) + \frac{\rho_1\sigma}{\sqrt{2\pi}}\exp\left(-\frac{a^2}{2\sigma^2}\right)\Phi\left(\frac{a-\rho_1 a}{\sigma\sqrt{1-\rho_1^2}}\right) \\
&= -\frac{\sigma}{\sqrt{2\pi}}\exp\left(-\frac{a^2}{2\sigma^2}\right) + (1+\rho_1)\frac{\sigma}{\sqrt{2\pi}}\exp\left(-\frac{a^2}{2\sigma^2}\right)\Phi\left(\frac{a-\rho_1 a}{\sigma\sqrt{1-\rho_1^2}}\right).
\end{aligned} \tag{5.10}$$

Step 2: Calculus of $\mathbb{E}(X_1 1_{(X_0,X_1)\in\Delta})$. Since (X_0, X_1) and $(-X_1, -X_0)$ have the same law, thus,

$$\mathbb{E}(X_1 1_{(X_0,X_1)\in\Delta}) = -f(-a,\rho_1).$$

Step 3: Calculus of $g(a,\rho_1) = \mathbb{P}((X_0,X_1)\in\Delta)$. This part is the most difficult. To calculate $g(a,\rho_1)$, as in the Step 1 we have

$$g(a,\rho_1) = \mathbb{E}1_{X_0\le a}\Phi\left(\frac{\rho_1 X_0 - a}{\sigma\sqrt{1-\rho_1^2}}\right).$$

However, It is also difficult to calculate the integral above explicitly. The trick consists to take the derivation of $g(\rho_1) := g(a,\rho_1)$ w.r.t ρ_1.

$$\begin{aligned}
g'(\rho_1) &= \frac{d}{d\rho_1}\mathbb{E}1_{X_0\le a}\Phi\left(\frac{\rho_1 X_0 - a}{\sigma\sqrt{1-\rho_1^2}}\right) \\
&= \mathbb{E}1_{X_0\le a}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\rho_1 X_0 - a}{\sigma\sqrt{1-\rho_1^2}}\right)^2\right)\frac{d}{d\rho_1}\left(\frac{\rho_1 X_0 - a}{\sigma\sqrt{1-\rho_1^2}}\right) \\
&= \mathbb{E}1_{X_0\le a}\frac{1}{\sqrt{2\pi}(1-\rho_1^2)}\frac{X_0 - a\rho_1}{\sigma\sqrt{1-\rho_1^2}}\exp\left(-\frac{1}{2}\left(\frac{\rho_1 X_0 - a}{\sigma\sqrt{1-\rho_1^2}}\right)^2\right) \\
&= \int_{-\infty}^{a}\frac{1}{\sqrt{2\pi}(1-\rho_1^2)}\frac{x - a\rho_1}{\sigma\sqrt{1-\rho_1^2}}\exp\left(-\frac{1}{2}\left(\frac{\rho_1 x - a}{\sigma\sqrt{1-\rho_1^2}}\right)^2\right)\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right)dx \\
&= \int_{-\infty}^{a}\frac{1}{2\pi(1-\rho_1^2)}\frac{x - a\rho_1}{\sigma^2\sqrt{1-\rho_1^2}}\exp\left(-\frac{1}{2}\left(\frac{x - a\rho_1}{\sigma\sqrt{1-\rho_1^2}}\right)^2\right)\exp\left(-\frac{a^2}{2\sigma^2}\right)dx \\
&= \exp\left(-\frac{a^2}{2\sigma^2}\right)\frac{1}{2\pi(1-\rho_1^2)}\sqrt{1-\rho_1^2}\int_{-\infty}^{\frac{a-a\rho_1}{\sigma\sqrt{1-\rho_1^2}}} u\exp\left(-\frac{u^2}{2}\right)du \\
&= -\frac{1}{2\pi\sqrt{1-\rho_1^2}}\exp\left(-\frac{a^2}{(1+\rho_1)\sigma^2}\right).
\end{aligned} \tag{5.11}$$

This completes the proof of Proposition 5.3. □

5.5 Proof of Theorem 5.1

Introduce the counting function $l : \mathbb{R}^+ \to \mathbb{R}^+$,

$$l(n) := \sum_{0}^{n-1} 1_{(X_k, X_{k+1}, \cdots, X_{k+d-1}) \in \Delta}, \;\; \Delta_k := \{(X_k, X_{k+1}, \cdots, X_{k+d-1}) \in \Delta\}.$$

For $\bar{D}_n(j)$, noting that

$$\bar{D}_n(j) = \frac{1}{l(n)} \sum_{k=1}^{l(n)} X_{\tau_k + j} = \frac{1}{l(n)} \sum_{k=0}^{n-1} 1_{\Delta_k} X_{k+j}$$

which converges a.s. to

$$\frac{1}{\mathbb{P}(\Delta_0)} \mathbb{E}\left(X_j 1_{\Delta_0}\right) = \mathbb{E}(X_j | \Delta_0)$$

by Birkhoff's ergodic theorem. Now for $D_n(j)$, consider the inverse of $l(\cdot)$ given by

$$l^{-1}(n) = \inf\{k; l(k) = n\}.$$

We have,

$$\begin{aligned} D_n(j) &= \frac{1}{n} \sum_{k=1}^{n} X_{\tau_k + j} = \frac{1}{n} \sum_{k=0}^{l^{-1}(n)-1} 1_{\Delta_k} X_{k+j} \\ &= \frac{l^{-1}(n)}{n} \frac{1}{l^{-1}(n)} \sum_{k=0}^{l^{-1}(n)-1} 1_{\Delta_k} X_{k+j} \end{aligned} \tag{5.12}$$

As the process is stationary and ergodic, so by the ergodic theorem of Birkhoff ,

$$\lim_{n\to\infty} \frac{l(n)}{n} = \lim_{n\to\infty} \frac{\sum_0^{n-1} 1_{\Delta_k}}{n} = \mathbb{P}(\Delta_0), \;\; a.s. \tag{5.13}$$

and

$$\lim_{n\to\infty} \frac{1}{l^{-1}(n)} \sum_{k=0}^{l^{-1}(n)-1} 1_{\Delta_k} X_{k+j} = \mathbb{E}(X_j 1_{\Delta_0}), \;\; a.s. \tag{5.14}$$

Then $l^{-1}(n)/n \to 1/\mathbb{P}(\Delta_0), a.s.$ and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_{\tau_k + j} = \frac{\mathbb{E}(X_j 1_{\Delta_0})}{\mathbb{P}(\Delta_0)} = \mathbb{E}(X_j | \Delta_0), \;\; a.s.$$

This completes the proof of Theorem 5.1.

5.6 Proof of Theorem 5.2

We begin with

Lemma 5.6 *(X_i) is a stationary and ergodic sequence, if*

$$\sum_{n=0}^{\infty} \rho^*(n) < +\infty, \tag{5.15}$$

then for any $f(X_i) \in L^2$ such that $\mathbb{E}(f) = 0$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} f(X_i) \xrightarrow{d} (\sigma B_t)_{t \geq 0}$$

in the space $\mathbb{D}(\mathbb{R}^+; \mathbb{R})$ of real càdlàg functions on $\mathbb{R}^+$ equipped with the topology of uniform convergence over compacts, where (B_t) is the standard Brown motion, and

$$\sigma^2 = \mathbb{E}f^2(X_0) + 2\sum_{i=1}^{\infty} \mathbb{E}f(X_0)f(X_i)$$

Proof. [Proof of Theorem 5.2]

The CLT for $\bar{D}(j)$ is quite easy. Indeed,

$$\begin{aligned}
\sqrt{l(n)}\left(\bar{D}_n(j) - \mathbb{E}(X_j|\Delta_0)\right) &= \frac{1}{\sqrt{l(n)}} \sum_{k=0}^{n-1} 1_{\Delta_k}\left(X_{k+j} - \mathbb{E}(X_{k+j}|\Delta_k)\right) \\
&= \sqrt{\frac{n}{l(n)}} \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} 1_{\Delta_k}\left(X_{k+j} - \mathbb{E}(X_{k+j}|\Delta_k)\right).
\end{aligned}$$

Applying Lemma 5.6 to the stationary and ergodic process $(X_k, \cdot, X_{k+m})_{k \geq 0}$ where $m \geq \max\{d-1, j\}$ (fixed), we have

$$\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} 1_{\Delta_k}\left(X_{k+j} - \mathbb{E}(X_j|\Delta_0)\right) \to \sigma B_1$$

in distribution, where

$$\sigma^2 = Var\left(1_{\Delta_0}\left(X_j - \mathbb{E}(X_j|\Delta_0)\right)\right) + 2\sum_{k=1}^{+\infty} Cov\left(1_{\Delta_0}\left(X_j - \mathbb{E}(X_j|\Delta_0)\right), 1_{\Delta_k}\left(X_{k+j} - \mathbb{E}(X_{k+j}|\Delta_k)\right)\right).$$

Since $n/l(n) \to 1/\mathbb{P}(\delta_0)$ a.s., we have by an elementary argument that

$$\sqrt{l(n)}\left(\bar{D}_n(j) - \mathbb{E}(X_j|\Delta_0)\right) \to \tilde{\sigma} B_1$$

in distribution, where $\tilde{\sigma}^2 = \sigma^2/\mathbb{P}(\Delta_0)$.

We now turn to the CLT of $D_n(j)$, which is technically more difficult. Note that

$$Y_n := \sqrt{n}\,(D_n(j) - \mathbb{E}(X_j|\Delta_0)) = \frac{1}{\sqrt{n}} \sum_{k=0}^{l^{-1}(n)} (1_{\Delta_k}\,(X_{k+j} - \mathbb{E}(X_j|\Delta_0))$$

and $l^{-1}(n)/n \to 1/\mathbb{P}(\Delta_0),\ a.s.$. For any $\delta > 0$ let

$$Z_n(t) := \frac{1}{\sqrt{n}} \sum_{k=0}^{[nt]} (1_{\Delta_k}\,(X_{k+j} - \mathbb{E}(X_j|\Delta_0))\,,\ A_n := \left\{ |\frac{l^{-1}(n)}{n} - \frac{1}{\mathbb{P}(\Delta_0)}| \le \delta \right\}.$$

For any bounded continuous function F on $\mathbb{R}$ with $|F| \le M$, as $\gamma \to F(\gamma(\cdot))$ is continuous from $\mathbb{D}(\mathbb{R}^+)$ to $\mathbb{D}(\mathbb{R}^+)$ w.r.t. the uniform over compacts convergence, we have by Lemma 5.6 that for $n \ge 1/\delta$

$$\begin{aligned} |\mathbb{E}F(Y_n) - \mathbb{E}F(Z_n(t_0))| &\le \mathbb{P}(A_n^c)M + \mathbb{E} \sup_{t:|t-t_0|\le 2\delta} |F(Z_n(t)) - F(Z_n(t_0))| \\ &\to \mathbb{E} \sup_{t:|t-t_0|\le 2\delta} |F(\tilde{\sigma}B_t) - F(\tilde{\sigma}B_{t_0})| \end{aligned}$$

where the last term is arbitrarily small once $\delta \to 0+$ again by Lemma 5.6. This completes the proof of the CLT for $D_n(j)$. □

5.7 Proof of Theorem 5.4

By Theorem 5.1 applied to (X_{kh}), we have a.s.

$$\lim_{n\to\infty} D_n(h,t) = \mathbb{E}(X_t|(X_0, X_h) \in \Delta).$$

By Proposition 5.3, the last quantity equals

$$\frac{\frac{\rho(t) + \rho(t-h)}{1+\rho(h)} f(a,\rho(h)) + \frac{\rho(t-h) - \rho(t)\rho(h)}{1+\rho(h)} \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2\sigma^2}\right)}{g(a,\rho(h))} \tag{5.16}$$

Note that as $h \to 0+$, $\rho(h) \to 1$ and then

$$f(a,\rho(h)),\ g(a,\rho(h)),\ \frac{\rho(t-h) - \rho(t)\rho(h)}{1+\rho(h)}$$

tend to zero totally, and

$$\frac{\rho(t)+\rho(t-h)}{1+\rho(h)} \to \rho(t).$$

For the behavior of $f(a,\rho(h))$ as $h \to 0+$, we calculate

$$\begin{aligned}
\frac{\partial f(a,\rho(h))}{\partial \rho(h)} &= \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2\sigma^2}\right) \Phi\left(\frac{a-\rho(h)a}{\sigma\sqrt{1-\rho^2(h)}}\right) \\
&\quad + (1+\rho(h))\frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2\sigma^2}\right) \frac{\partial}{\partial \rho(h)} \Phi\left(\frac{a-\rho(h)a}{\sigma\sqrt{1-\rho^2(h)}}\right) \\
&= \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2\sigma^2}\right) \left\{ \Phi\left(\frac{a-\rho(h)a}{\sigma\sqrt{1-\rho^2(h)}}\right) \right. \\
&\quad \left. - \frac{a}{\sigma\sqrt{2\pi(1-\rho^2(h))}} \exp\left(-\frac{a^2(1-\rho(h))}{2\sigma^2(1+\rho(h))}\right) \right\}
\end{aligned} \tag{5.17}$$

Thus, when $h \to 0+$,

$$\frac{\partial f(a,\rho(h))}{\partial \rho(h)} \sim \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2\sigma^2}\right) \left\{ \frac{1}{2} - \frac{a}{2\sigma\sqrt{\pi(1-\rho(h))}} \right\}, \tag{5.18}$$

$$\frac{\partial g(a,\rho(h))}{\partial \rho(h)} \sim -\frac{1}{2\sqrt{2}\pi\sqrt{1-\rho(h)}} \exp\left(-\frac{a^2}{2\sigma^2}\right) \tag{5.19}$$

So by the Hospital's criterion,

$$\begin{aligned}
\lim_{h\to 0} \frac{f(a,\rho(h))}{g(a,\rho(h))} &= \lim_{\rho(h)\to 1} \frac{\frac{\partial f(a,\rho(h))}{\partial \rho(h)}}{\frac{\partial g(a,\rho(h))}{\partial \rho(h)}} \\
&= \lim_{\rho(h)\to 1} \frac{\frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2\sigma^2}\right) \left(\frac{1}{2} - \frac{a}{2\sigma\sqrt{\pi(1-\rho(h))}} \right)}{-\frac{1}{2\sqrt{2}\pi\sqrt{1-\rho(h)}} \exp\left(-\frac{a^2}{2\sigma^2}\right)} = a
\end{aligned} \tag{5.20}$$

Furthermore again by Hospital's criterion,

$$\lim_{\rho(h)\to 1-} \frac{g(a,\rho(h))}{\sqrt{1-\rho(h)}} = \lim_{\rho\to 1-} 2u(\rho)\sqrt{1-\rho} = \frac{1}{2\sqrt{2}\pi} \exp\left(-\frac{a^2}{2\sigma^2}\right) =: c \tag{5.21}$$

where $u(\rho)$ is given in Proposition 5.3.

Now we separate our discussion into two cases.

Part (a). $\Gamma := \int_{\mathbb{R}} x^2\mu(dx) < +\infty$. In that case $\rho \in C^2(\mathbb{R})$ and

$$\rho'(0) = 0,\ \sigma^2\rho''(0) = -\Gamma \in (0, +\infty).$$

and

$$\rho(h) = 1 + \frac{1}{2}\rho''(0)h^2 + o(h^2)$$
$$\rho(t-h) - \rho(t)\rho(h) = -\rho'(t)h + o(h)$$

we have by (5.21),

$$\begin{aligned}\lim_{h\downarrow 0} \frac{\rho(t-h) - \rho(t)\rho(h)}{g(a, \rho(h))} &= \lim_{h\downarrow 0} \frac{-\rho'(t)h + o(h)}{c\sqrt{-\frac{1}{2}\rho''(0)h^2 + o(h^2)}} \\ &= -2\pi \exp\left(\frac{a^2}{2\sigma^2}\right)\rho'(t)\frac{1}{\sqrt{-\rho''(0)}}.\end{aligned} \tag{5.22}$$

Substituting it together with (5.20) into (5.16), we obtain

$$\lim_{h\to 0} \mathbb{E}(X_t|(X_0, X_h) \in \Delta) = \rho(t)a - \sqrt{\frac{\pi}{2}}\frac{\rho'(t)\sigma}{\sqrt{-\rho''(0)}}$$

the desired result in (a).

Part (b). $\Gamma = +\infty$. By (5.16), (5.20) and (5.21), it remains to prove that

$$\lim_{h\to 0+} \frac{\rho(t-h) - \rho(t)\rho(h)}{\sqrt{1-\rho(h)}} = \lim_{h\to 0+} \frac{\rho(t-h) - \rho(t) + \rho(t)(1-\rho(h))}{\sqrt{1-\rho(h)}} = 0. \tag{5.23}$$

Let $\mu_n(dx) := 1_{[-n,n]}\mu(dx)$. Then

$$1 - \rho(h) = \int_{\mathbb{R}} (1 - \cos hx)\frac{1}{\sigma^2}\mu(dx) \geq \int_{\mathbb{R}} (1 - \cos hx)\frac{1}{\sigma^2}\mu_n(dx) = \frac{1}{2}\frac{\Gamma_n}{\sigma^2}h^2 + o(h^2).$$

where $\Gamma_n := \int_{[-n,n]} x^2\mu(dx)$. Consequently

$$\limsup_{h\to 0+} \frac{|\rho(t-h) - \rho(t)|}{\sqrt{1-\rho(h)}} \leq \lim_{h\to 0+} \frac{|\rho'(t)h + o(h)|}{\sqrt{\frac{1}{2}\frac{\Gamma_n}{\sigma^2}h^2 + o(h^2)}} = \sigma|\rho'(t)|\sqrt{\frac{2}{\Gamma_n}}$$

where the desired result follows by letting $n \to +\infty$.

Remarks 5.7 *By the proof above (see (5.23)), the asymptotic consistence of the RDEs $D_n(h,t)$ and $\bar{D}_n(h,t)$ is equivalent to*

$$\lim_{h\to 0}\frac{\rho(t-h)-\rho(t)}{\sqrt{1-\rho(h)}}=0$$

(this is without the assumption that $\rho'(t)$ exists for $t>0$!).

5.8 An example

Consider the case of an harmonic oscillation excited by a white noise which is given by the following Itô stochastic differential equation,

$$\begin{cases} dX_t = V_t dt \\ dV_t = \varsigma dW_t - (2\kappa\omega_0 V_t + {\omega_0}^2 X_t)dt \end{cases} \tag{5.24}$$

where ω_0 is the natural pulsation, and κ is the damping coefficient ($\omega_0>0,\ \kappa>0$).

Set $Y_t = \begin{pmatrix} X_t \\ V_t \end{pmatrix}$, it is easy to see that (Y_t) is a Markov process. Rewrite the equation:

$$d\begin{pmatrix} X_t \\ V_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega_0^2 & -2\kappa\omega_0 \end{pmatrix}\begin{pmatrix} X_t \\ V_t \end{pmatrix}dt + \begin{pmatrix} 0 \\ \varsigma \end{pmatrix}dW_t \tag{5.25}$$

the solution to this equation is explicitly given by

$$Y_t = e^{At}(Y_0 + \int_0^t e^{-As}BdW_s) \tag{5.26}$$

where

$$A = \begin{pmatrix} 0 & 1 \\ -\omega_0^2 & -2\kappa\omega_0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ \varsigma \end{pmatrix}.$$

By [13], the unique stationary distribution is given by $\mathcal{N}(0,\Sigma)$ where

$$\Sigma := \begin{pmatrix} var(X_0) & 0 \\ 0 & var(V_0) \end{pmatrix} = \begin{pmatrix} \dfrac{\varsigma^2}{4\kappa{\omega_0}^3} & 0 \\ 0 & \dfrac{\varsigma^2}{4\kappa\omega_0} \end{pmatrix}. \tag{5.27}$$

Below we take the initial variable Y_0 independent of $(W_t)_{t\geq 0}$ with law $\mathcal{N}(0,\Sigma)$. By the Markov property of (Y_t),

$$\rho^*(t) = \sup_u \rho(\sigma(X_s, s\leq u); \sigma(X_s, s\geq u+t)) = \rho(\sigma(Y_0), \sigma(Y_t))$$

By [14, Remarks 5.3] we have

$$(\rho^*(t))^2 = \lambda_{max}(\Sigma^{-1/2} e^{A^* t} \Sigma e^{At} \Sigma^{-1/2}) \tag{5.28}$$

where A^* is the transposition of A, λ_{max} denote the maximal eigenvalue of a matrix.

We assume at first that $\kappa < 1$. The two eigenvalues of A are given by $\lambda_{1,2} = \omega_0(-\kappa \pm i\sqrt{1-\kappa^2})$, then $A = Q\Lambda Q^{-1}$, where

$$Q = \begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix}, \quad \Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \quad Q^{-1} = \frac{1}{\lambda_2 - \lambda_1} \begin{pmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{pmatrix}$$

(note that they are complex-valued). Thus letting $\|A\| := \sup_{x \in \mathbb{C}^n : |x|=1} |Ax|$ be the norm of complex matrix A, we have by (5.28)

$$\begin{aligned} \rho^*(t) &= \|\Sigma^{1/2} e^{At} \Sigma^{-1/2}\| = \|\Sigma^{1/2} Q e^{t\Lambda} Q^{-1} \Sigma^{-1/2}\| \\ &\le \|e^{t\Lambda}\| C = e^{-\omega_0 \kappa t} C \end{aligned}$$

where

$$C = \|\Sigma^{1/2} Q\| \cdot \|Q^{-1} \Sigma^{-1/2}\|.$$

The same calculus works for $\kappa > 1$ (except $\lambda_1, \lambda_2, Q, \Lambda$ are all real-valued).

In the case where $\kappa = 1$, A is not diagonizable. But by calculting first explicitly $\Sigma^{-1/2} e^{At} \Sigma e^{A^* t} \Sigma^{-1/2}$ for $\kappa \neq 1$ and next let $\kappa \to 1$, we shall obtain

$$\rho^*(t) \le e^{-\omega_0 \kappa t}(c_0 + c_1 t)$$

for some constants $c_0, c_1 > 0$.

Hence for any $\kappa > 0$, for each mesh h fixed, the CLT in Theorem 5.2 is applicable.

We now turn to the consistency. In fact for all $t \ge 0$, $\rho(t) = \dfrac{Cov(X_0, X_t)}{Var(X_0)}$, by (5.26) and (5.27) we have $Cov(X_t, X_0) = \left(e^{At} Cov(Y_0, Y_0)\right)_{11} = \left(e^{At}\Sigma\right)_{11}$. Thus by the diagolization of A above we get that for $\kappa \neq 1$,

$$\rho(t) = \frac{1}{\lambda_2 - \lambda_1}\left[(\lambda_2 e^{\lambda_1 t} - \lambda_1 e^{\lambda_2 t}) + \omega_0^2 (e^{\lambda_2 t} - e^{\lambda_1 t})\right] \tag{5.29}$$

Recall that the expression above is for $t \ge 0$ and for $t < 0$, $\rho(t) := \rho(-t)$. We now separate our discussion into three cases:

Case 1. $\kappa > 1$**.** We have $\rho'(0+) = \omega_0^2 > 0$. As a pair function on $\mathbb{R}$, ρ is not derivable at 0, hence the RDEs $D_n(h,t), \bar{D}_n(h,t)$ are consistent estimators of $a\rho(t)$ by Theorem 5.4.

Case 2. $\kappa \in (0,1)$**.** In that case,

$$\rho(t) = \frac{e^{-\omega_0 \kappa t}}{\sqrt{1-\kappa^2}}\left(\kappa \sin(\omega_0 t\sqrt{1-\kappa^2}) + \sqrt{1-\kappa^2}\cos(\omega_0 t\sqrt{1-\kappa^2})\right)$$

then $\rho'(0+) = \omega_0^2 > 0$. Hence the RDEs $D_n(h,t), \bar{D}_n(h,t)$ are consistent estimators of $a\rho(t)$ by Theorem 5.4.

Case 3. $\kappa = 1$ **(resonance case).** We can take limit $\kappa \to 1-$ in (5.29), and get immediately that for $t \geq 0$,

$$\rho(t) = (\omega_0\kappa t + 1)e^{-\omega_0\kappa t}.$$

We find that $\rho'(0+) = 0$ and $\rho''(0+) = -(\omega_0\kappa)^4$. Thus ρ is twice derivable at 0. By Theorem 5.4, we have a.s.

$$\lim_{h=t/m\to 0}\lim_{n\to\infty} D_n(h,t) = \rho(t)a - \sqrt{\frac{\pi Var(X_0)}{-2\rho''(0)}}\rho'(t) = \rho(t)a + \sqrt{\frac{\pi\varsigma^2}{8\kappa{\omega_0}^3}}te^{-\omega_0\kappa t}$$

and same for $\bar{D}_n(h,t)$. In other words in this resonance case $D_n(h,t), \bar{D}_n(h,t)$ have an asymptotic bias $\sqrt{\frac{\pi\varsigma^2}{8\kappa{\omega_0}^3}}te^{-\omega_0\kappa t}$.

Bibliography

[1] Bernard, P. (1988) Identification de grandes structures: une remarque sur la méthode du décrément aléatoire. *Journal de Mécanique théorique et appliquée*, **7**, 3, 269-280.

[2] Billingsley, P. (1968) *Convergence of Probability Measures*, John Wiley and Sons Inc.

[3] Caughey, T.K. and Stumpf, H.J. (1961) Transient response of dynamical system under random excitation. *Jal of Applied Mechanics*, **4**, 563-566.

[4] Chung, K.L. (1974)*A course in probability theory*. Second Edition, Academic Press New York and London.

[5] Cole, H.A. (1973) On-line failure Detection and damping Measurement of Aerospace Structures by Random Decrement Signatures. NASA CR, 2205.

[6] Hall, P. and Heyde, C.C. (1980) *Martingale limit theory and its application*, Academic Press INC, New York.

[7] Ibragimov, I.A. and Linnik, Y.V. (1971) *Independent and stationary sequences of variables random* , Ed. by J.F.C. Kingman, Wolters-Noordhoff, Groningen.

[8] Ibrahim, S.R. and Mikulcik, E.C. (1977) A Method for ther Direct Identification of Vibration Parameters from the Free Response. *The Shock and Vibration Bulletin*, **47**, 183-198.

[9] Kolmogorov A. N. and Rozanov Yu. A. (1960) On strong mixing conditions for stationary Gaussian processes. *Theory of Probability and Its Applications*, **5**, 204-208.

[10] Kac,M. and Slepian, D. (1959) Large excursions of Gaussian processes. *Annals of Math Stat*, **30**, 1215-1228.

[11] Meyn, S.P. and Tweedie, R.L. (1993)*Markov chains and stochastic stability*, Springer-Verlag, London.

[12] Sarmanov, O.V. (1958) Maximal coefficient of correlation (nonsymmetric case), *Doklady Akademii Nauk SSSR*, **121**, no. 1 (in Russian).

[13] Wu, L.M. (2001) Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems, *Stochastic Process. Appl.*, **91**, 205-238.

[14] Wu, L.M. (2005) A Φ-entropy contraction inequality for Gaussian vectors and applications to Gaussian Markov processes. Preprint 2005.

RESUME

Cette thèse est consacrée a l'étude de deux thèmes : les grandes déviations pour les estimateurs à noyau de la densité f_n^* des processus stochastiques stationnaires et l'estimateur de décrément aléatoire (EDA) pour des processus gaussiens stationnaires.

Le premier thème est la partie principale de cette thèse, constituée des quatre premiers chapitres. Dans le chapitre 1, on établit le w*-PGD (principe de grandes déviations) de f_n^* et une inégalité de concentration dans le cas i.i.d. On démontre dans le chapitre 2 la convergence exponentielle de f_n^* dans $L^1(\mathbb{R}^d)$ et une inégalité de concentration pour des suites ϕ-mélangeantes, en se basant sur une inégalité de transport de Rio. Les chapitres 3 et 4 constituent le coeur de cette thèse : on établit (i) le PGD de f_n^* pour la topologie faible $\sigma(L^1, L^\infty)$; (ii) le w*-PGD de f_n^* dans L^1 pour la topologie forte $\|\cdot\|_1$; (iii) l'estimation de grandes déviations pour l'erreur $D_n^* = \|f_n^*(x) - f(x)\|_1$ et (iv) l'optimalité asymptotique de f_n^* au sens de Bahadur. Ces résultats sont prouvés dans le chapitre 3 pour des processus de Markov uniformément ergodiques et dans le chapitre 4 pour des procesus de Markov réversibles uniformément intégrables.

Le dernier chapitre est consacré au second thème. On démontre la loi des grands nombres et le théorème de limite centrale pour l'EDA à temps discret et on établit pour la première fois l'expression explicite du biais de l'EDA à temps continu.

ABSTRACT

This thesis is devoted to the studies of two themes: large deviations of the kernel density estimator for stationary stochastic processes and random decrement estimator (RDE) for stationary gaussian processes.

The first theme is the main part of this thesis and contains four chapters. In chapter 1, we establish the w*-LDP (large deviation principle) of f_n^* and a concentration inequality in the i.i.d. case. In chapter 2, we prove the exponential convergence of f_n^* dans $L^1(\mathbb{R}^d)$ and a concentration inequality for the ϕ-mixing processes, by using a transportation inequality of Rio. Chapter 3 and chapter 4 are the core of this thesis. For the first time in the dependent case, we establish (i) the LDP of f_n^* for the weak topology $\sigma(L^1, L^\infty)$; (ii) the weak*-LDP of f_n^* in L^1 for the strong topology $\|\cdot\|_1$; (iii) large deviations estimations for $D_n^* = \|f_n^*(x) - f(x)\|_1$ and (iv) aymptotic optimality of f_n^* in the sense of Bahadur. These results are established in chapter 3 for uniformly ergodic Markov processes, and for uniformly integrable reversible Markov processes in chapter 4.

The last chapter is devoted to the second theme. We prove the law of large number and central limit theorem for RDE in discret time case and give the explicit bias of the RDE in continuous time case.